



SeqOne platform

Instruções de utilização

Válido a partir da SeqOne Platform v2.1

1. Introdução	7
2. Contactar o fabricante e o representante na Suíça	8
3. Vigilância dos dispositivos médicos	8
4. Utilização prevista da plataforma	8
4.1. Utilização prevista	8
4.2. Indicações de utilização	9
4.3. Descrição	9
4.4. População de pacientes	10
4.5. Utilizadores previstos	10
4.6. Caraterísticas	10
4.6.1. Desempenho analítico	12
4.6.1.1. Módulos bioinformáticos	13
4.6.1.2. Módulos de Anotação	21
4.6.2. Desempenho clínico	23
5. Precauções	24
5.1. Utilização conforme	24
5.2. Avisos	24
5.3. Limitações	26
5.4. Avarias	26
6. Requisitos mínimos	27
6.1. Hardware	27
6.2. Rede	27
7. Autenticação	28
7.1. Iniciar sessão	28
7.2. Terminar sessão	29
7.3. Página inicial	29
7.4. Gestão de contas de utilizador	30
7.4.1. Gestão de "My Account"	31
7.4.2. Gerir as definições do utilizador - definições	31
Entity	31
Users	31
Worksets	31
Manifest	31
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. Gestão administrativa - apresentação da interface	34
7.4.4. Alternando entre os modos Main e Preview	37
8. Descrição das etapas operacionais	38
9. Criação de projetos	39
9.1. Criar um projeto	39
9.2. Gestão de projetos	40
9.3. Eliminar um projeto	40
9.4. Dados de ADN (sequenciação de ADN, sequenciação de ADNct)	40
9.5. Dados de ARN (Sequenciação de ARN)	42
9.6. Definições do projeto	43
9.6.1. Configurar métricas QC padrão do projeto	43
9.6.2. Definir o perfil de filtro predefinido do projeto	43
10. Carregamento de amostras	44
10.1. Carregamento de amostras	44
A partir de um Computador	44
Via SDS (SeqOne Data Sync)	44
Estado de carregamento de amostras	44
Capacidade de carregamento	45
10.2. Formato do ficheiro de entrada	45
10.2.1. Ficheiros de análise primária	45
10.2.2. Ficheiros de análise secundários	46
10.3. Metadados da amostra	48
10.4. Editar uma amostra	49
10.5. Eliminar uma amostra	49
11. Análise	50
11.1. Lançamento de uma nova análise	50
11.1.1. Escolher o workset de análise	51
Worksets de análise disponíveis a partir de ficheiros FASTQ	51
GermlineVar	52
GermlineVarLR	52
GermlineFamily	53
GermVar Family	53
CNVCapture	54
CNVExome	54
SomaVar	54
SomaCNVCapture	55
SomaRNA	55
SomaMSI	55

SomaHRD	56
SomaVar LF	56
SomaCGP	57
SomaHemato	57
SomaLBx	58
SomaMethyl	58
Worksets de análise disponíveis a partir de ficheiros de análises secundárias	58
GermVar Tertiary	59
GermVar Tertiary Family	59
SomaVar Tertiary	59
11.1.2. Escolher o modo de processamento de UMI	60
Apresentação dos modos de processamento de UMI	60
Processo de tratamento de UMI	61
Compatibilidade dos worksets de análise	61
11.2. Estado da análise	62
11.3. Qualidade da amostra	62
12. Controlo de qualidade dos dados	64
12.1. Métricas e relatórios QC	64
12.1.1. Métricas de QC e controlos de QC	64
12.1.2. Relatório QC	64
12.1.3. Relatório de monitorização da identidade 'ID-Check'.	68
12.1.4. Relatório técnico	69
12.1.5. Relatório UPD	70
Pontuação UPD	70
Gráfico de frequência do alelo B (BAF) do relatório UPD	71
13. Interpretação dos resultados da análise	72
13.1. Resultados da análise	72
13.1.1. Gestão do fluxo de trabalho de interpretação das análises	72
13.1.1.1. Atribuição de análises para interpretação	72
13.1.1.2. Bloqueio de análises	73
13.1.1.3. Painel lateral para gestão da análise	74
Visualização modular do Workflow	75
Visualização modular do Patient	76
Visualização modular do HPOs	77
Visualização modular do Panel	79
13.1.1.4. Painel lateral para gerir as avaliações	80
13.1.2. Visão geral dos resultados da análise no separador Dashboard	81
13.1.3. Cobertura pormenorizada por exão.	82
Metodologia	82
Visualização e interpretação	83

13.1.4. Variantes pequenas	86
Metodologia	86
Visualização e interpretação	90
Ordenação de variantes	104
Configuração das colunas da tabela de variantes	106
Editar os metadados da amostra a partir do painel de análise	106
Variantes de filtragem	108
Filtros Genes & Variants	108
Secção Samples & Quality	108
Secção Frequencies	110
Secção Genetics	111
Secção Databases	112
Secção Transmission & Phenotypes	113
Secção Classification	116
Secção Scores	117
Selecionar uma doença	119
Utilizar um perfil de filtros predefinido	120
Editar e guardar perfis de filtros	120
Perfis de filtros favoritos	120
Combinação de perfis de filtro	121
Perfis de filtros SeqOne	121
Função "Mark as viewed"	122
Selecionar variantes	122
Avaliação de variantes	123
Página de variantes	126
Geração de um relatório clínico	135
13.1.5. CNV	136
13.1.5.1. CNV em painéis de genes	136
Metodologia	136
Visualização e interpretação	137
13.1.5.2. CNV na análise SomaCGP e SomaLBx	142
Metodologia	142
Visualização e interpretação	142
13.1.5.3. CNV nos exomas e genomas	143
Metodologia para o exoma	143
Metodologia para o genoma	144
Visualização e interpretação	144
Tabela de resultados CNV	145
Página pormenorizada do CNV	149
13.1.6. Visualizador de variantes grandes para CNVs e SVs (LVV)	154

Tabela de resultados de CNVs & SVs	154
Classificação, Filtragem, Visualização e Interpretação	156
Painel lateral CNV/SV (Details Drawer)	157
13.1.7. Assinaturas na análise SomaCGP	158
Metodologia para calcular a pontuação MSI	158
Metodologia para calcular a pontuação TMB	158
Visualização e interpretação	159
13.1.8. Repetições curtas em tandem (STRs)	161
Metodologia	161
Visualização e interpretação	161
13.1.9. Fusão de Genes	163
13.1.9.1. Fusões detetadas a partir do ADN	163
Metodologia	163
Visualização e interpretação	163
13.1.9.2. Fusões detetadas a partir de ARN	165
Metodologia	165
Visualização e interpretação	166
13.1.10. Clonalidade	169
13.1.11. Variantes de splicing	170
Metodologia	170
Visualização e interpretação	170
13.1.12. Estado MSI/MSS	172
Metodologia	172
Visualização e interpretação	172
13.1.13. Estado HRD/HRP	174
Metodologia	174
Visualização e interpretação	176
13.1.14. Dados de metilação para SomaMethyl	182
14. Relatório	185
14.1.1. Ficheiros de saída	185
15. Resolução de problemas e mensagens de erro	186
16. Garantia	188
17. Aviso legal	188
17.1. Termos de utilização	188
17.2. Limitações de responsabilidade	188
17.3. Direitos de autor e marcas registadas	188

1. Introdução

Aconselha-se o utilizador a ler as instruções cuidadosamente antes de utilizar o *software*. Uma versão em papel destas instruções pode ser fornecida mediante pedido à equipa de suporte da SeqOne (support@seqone.com). Para obter os melhores resultados ao utilizar a plataforma,

leia atentamente todas as precauções destacadas com o símbolo . Os vários símbolos regulamentares e outros símbolos utilizados ao longo deste guia e do software são os seguintes:

Símbolo	Significado
	Fabricante
	Indica o representante autorizado estabelecido na Suíça.
	Data de fabrico (data de lançamento do software)
	Dispositivo médico de diagnóstico in vitro.
	Identificador único do dispositivo
	Referência do catálogo do fabricante
	Cumprir os requisitos do Regulamento Europeu (UE) 2017/746 Marcação CE com o número de identificação do organismo notificado
	Consultar as instruções de utilização
	Aviso Este símbolo de segurança indica que são necessários cuidados especiais aquando da utilização do aparelho.

A etiqueta (incluindo informações sobre a versão do software SeqOne e detalhes de contacto) e este documento que descreve as instruções de utilização (em várias línguas) estão disponíveis na secção "About".

2. Contactar o fabricante e o representante na Suíça



SeqOne S.A.S

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

22 rue Durand
34000 Montpellier
FRANÇA

Apoio ao utilizador :

support@seqone.com



Arazy Group Swiss GmbH

Bruderholzallee 53

4059 Basileia, Suíça

swiss.ar@arazygroup.com

3. Vigilância dos dispositivos médicos

Se ocorrer ou for provável que ocorra um incidente grave durante a utilização do dispositivo, pede-se aos utilizadores que comuniquem o incidente, sem demora injustificada, ao fabricante e à autoridade competente do país onde o incidente ocorreu ou é provável que ocorra.

Os regulamentos aplicáveis e a lista de autoridades competentes na UE estão disponíveis no seguinte endereço:

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

IVD

CE 0459

4. Utilização prevista da plataforma

4.1. Utilização prevista

A Plataforma SeqOne é um software que funciona como dispositivo médico de diagnóstico *in vitro*, concebido para apoiar os profissionais de saúde na análise e interpretação de dados genómicos brutos e de sequenciação de pacientes submetidos a análises NGS ou *arrays* CGH. Ao fornecer informações qualitativas e acionáveis, a plataforma funciona como um sistema de apoio à decisão, auxiliando os profissionais de saúde a tomar decisões clínicas.

4.2. Indicações de utilização

A Plataforma SeqOne suporta análises secundárias e terciárias a partir de dados de sequenciação em bruto, bem como análises terciárias isoladas. As suas capacidades analíticas, independentes da patologia, têm aplicações em diversos domínios médicos, incluindo a oncologia e as doenças raras.

4.3. Descrição

A Plataforma SeqOne é um software que funciona como dispositivo médico de diagnóstico *in vitro*, integrando um conjunto de aplicações que funcionam em sincronia para fornecer aos utilizadores especializados um poderoso conjunto de ferramentas para a interpretação de dados de sequenciação genómica complexos..

A Plataforma SeqOne integra informações estabelecidas a partir de uma vasta gama de **fontes externas**, incluindo bases de dados públicas, literatura científica e bases de conhecimento proprietárias. Estas fontes de dados externas fornecem contexto e significado às variantes genómicas identificadas.

Os ficheiros de entrada para várias aplicações dentro da plataforma são processados utilizando uma combinação personalizada de módulos de bioinformática e anotação, garantindo que cada análise seja otimizada para os seus requisitos específicos.

Os **módulos bioinformáticos** efectuam análises secundárias utilizando métodos computacionais, como alinhamento e a detecção, para identificar variantes genéticas em dados de sequenciação de ADN e ARN, comparando-os contra um genoma de referência (hg19 ou hg38). Os dados, que podem ser usados tanto para análise germinativa quanto somática, podem ser gerados a partir do sequenciamento do genoma completo, do sequenciamento do exoma ou de painéis de genes específicos.

Os **módulos de anotação** efectuam uma análise terciária para anotar as variantes em detalhe, como o gene e a sua função conhecida, a localização da variante no gene e quaisquer publicações ou relatórios clínicos que mencionem o gene ou a variante. Esta anotação determinará quais as regras de classificação ACMG e AMP/CAP/ASCO aplicadas a cada variante, permitindo chegar a um veredito sobre a patogenicidade. A anotação também é efetuada através do DiagAI, um conjunto de ferramentas baseadas em inteligência artificial, que ajuda a analisar variantes genéticas para identificar rapidamente potenciais variantes causais.

Esta arquitetura modular permite que a Plataforma SeqOne seja altamente adaptável a diferentes tipos de análises e questões em investigação. Os ficheiros de entrada para várias aplicações na plataforma são processados utilizando uma combinação personalizada destes módulos de Bioinformática e Anotações, garantindo que cada análise é otimizada para os seus requisitos específicos. Ao integrar estas diversas fontes de dados e ferramentas de análise, a Plataforma SeqOne permite que os utilizadores especializados obtenham conhecimentos

profundos sobre o complexo panorama dos dados genómicos e as suas implicações para a saúde e na doença humana.

A plataforma funciona como uma solução SaaS baseada na nuvem, com todas as funcionalidades acessíveis mediante um navegador de *internet*. Inclui uma ferramenta opcional, SeqOne Flow, que pode ser implementada na nuvem AWS ou nos servidores do cliente para facilitar a transferência de dados e a integração do fluxo de trabalho.

Por razões de segurança, todos os dados são encriptados em repouso, utilizando a encriptação AES-256, e o tráfego entre os utilizadores e a plataforma é protegido utilizando certificados SSL.

4.4. População de pacientes

A solução é adequada para qualquer paciente que necessite de uma análise genómica. Os pacientes alvo podem ser pessoas com cancro, doenças raras, doenças hereditárias, uma predisposição genética para o cancro ou pessoas saudáveis.

4.5. Utilizadores previstos

A Plataforma SeqOne foi concebida para profissionais de saúde responsáveis pela análise genómica clínica de rotina. Isto inclui a análise e interpretação de ficheiros brutos gerados pelo sequenciador a partir de amostras biológicas sequenciadas. Os utilizadores alvo da plataforma são geneticistas, bioinformáticos, biólogos, engenheiros e técnicos de plataformas NGS.

4.6. Caraterísticas

Cada aplicação (workset) é composta por vários **módulos de bioinformática e anotação**, cada um responsável por tarefas analíticas específicas. Neste capítulo, os desempenhos são avaliados ao nível do módulo para garantir uma avaliação detalhada da precisão e fiabilidade de cada componente.

Os analitos detetados pela SeqOne Platform enquadram-se em duas categorias de alterações genómicas:

- Variantes genéticas, tais como SNVs, indels, MNVs, CNVs, UPD, fusões genéticas, variantes estruturais, variantes splicing, inserções Alu e FLT3-ITD (duplicação em tandem interna).
- Assinaturas ou biomarcadores complexos, tais como o estado do IGHV, a instabilidade de sítios de microssatélites (MSI), a Tumor Mutational Burden (TMB), a metilação ou os defeitos na reparação de quebras de cadeia dupla por recombinação homóloga (HRD, Homologous Recombination Deficiency).

Módulo de bioinformática	Workset	Tipo de analito
Detecção de CNV Exoma Germline	CNVExome, GermVar, GermVar Family	Variante genética
Detecção de CNV Painel Germline	CNVCapture	Variante genética
SNV/Indel Germline	GermlineVar, GermlineFamily	Variante genética
SV da linha Germline	GermlineVar, GermlineFamily	Variante genética
UPD	GermVar, GermVar Family, GermVar Tertiary, GermVar Tertiary Family	Variante genética
Inserções Alu	GermlineVar, GermlineFamily, SomaVar, Somavar LF, SomaCGP, SomaHemato, GermVar, GermVar Family	Variante genética
MNV	GermlineVar, GermlineFamily, GermVar, GermVar Family, SomaVar Tertiary, GermVar Family Tertiary, SomaVar LF, SomaCGP, SomaHemato, SomaLBx, SomaVar Tertiary	Variante genética
SNV/Indel Somatic	SomaVar, SomaHemato, SomaRNA, SomaCGP	Variante genética
SNV/Indel Somatic LF	SomaVar LF	Variante genética
CNV Painéis somáticos de grande dimensão	SomaCGP	Variante genética
Índice de metilação	SomaMethyl	Assinatura
Fusão de ADN somático	SomaCGP, SomaVar, SomaVar LF, SomaHemato	Variante genética
Fusão de ADN somático LBx	SomaLBx	Variante genética
TMB	SomaCGP	Assinatura
MSI	SomaCGP, SomaMSI	Assinatura
Probabilidade GI (GiP)	SomaHRD	Assinatura
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
RNA de fusão	SomaRNA	Variante genética
RNA de splicing	SomaRNA	Variante genética
CNV Panel pequeno somático	SomaCNVCapture	Variante genética
Estatuto IGHV	SomaHemato	Assinatura
Subconjuntos IGHV (subsets)	SomaHemato	Assinatura
CFH	GermlineVarLR	Variante genética
FLT3	SomaVar, SomaHemato	Variante genética

4.6.1. Desempenho analítico

O desempenho indicado neste manual de instruções corresponde ao desempenho das versões IVD mais recentes dos worksets, que estão listadas na tabela abaixo:

Nome do workset	Versão
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

- **Detecção de CNV no exoma germline e no painel germline**

Tipo de variante	Versão do genoma	Controlos positivos	Verdadeiros positivos	Falsos negativos	Critérios de aceitação	Sensibilidade
Detecção de CNV exoma (GermlineVar)	GRCh37	8	8	0	100%	100%
Detecção de CNV exoma (GermVar)	GRCh37	8	8	0	100%	100%
Detecção de CNV Panel Germline	GRCh37	15	15	0	100%	100%
Detecção de CNV Panel Germline	GRCh38	11	11	0	100%	100%

- **SNV / Indel Germline - Dados do painel**

Versão do genoma	Tipo de variante	Critérios de aceitação	Sensibilidade antes do filtro	Sensibilidade após o filtro
GRCh37	SNV	100%	98%	100%
GRCh38	SNV	100%	98%	100%

Versão do genoma	Tipo de variante	Critérios de aceitação	Precisão antes do filtro	Precisão após o filtro
GRCh37	SNV	100%	90%	100%
GRCh38	SNV	100%	90%	100%

- **SNV / Indel Germline - Dados do exoma**

Versão do genoma	Tipo de variante	Critérios de aceitação*	Sensibilidade antes do filtro	Sensibilidade após o 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 a detecção de variantes localizadas nas regiões de alta confiança de NA12878 e NA24385.

Versão do genoma	Tipo de variante	Critérios de aceitação*	Precisão antes do filtro	Precisão após o 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 a detecção de variantes localizadas nas regiões de alta confiança de NA12878 e NA24385.

- **SNV / Indel Germline - Repetibilidade e Reprodutibilidade**

- **Repetibilidade**

Tipo de variante	Sensibilidade	Precisão
SNV	98,59% ± 0%	95,92% ± 0%
indel	87,75% ± 0%	81,63% ± 0%
(all)	98,38% ± 0%	95,63 ± 0%

- **Reprodutibilidade**

Tipo de variante	Sensibilidade	Precisão
SNV	98,57% ± 0,03%	95,93% ± 0,08%
indel	86,58% ± 0,42%	78,97% ± 0,96%
total	98,34% ± 0,03%	95,58 ± 0,09%

- **Variantes mosaicas Germline**

- **Sensibilidade SNV**

Versão do genoma	Intervalos VAF	N.º de variantes verdadeiras	Crítérios de aceitação	Sensibilidade e 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%

- **Sensibilidade Indels**

Versão do genoma	Intervalos VAF	N.º de variantes verdadeiras	Crítérios de aceitação	Sensibilidade e 100x
GRCh38	[0,05,0,1[	87	40%	43,7%
GRCh38	[0,1,0,2[	14	70%	78,6%
GRCh38	[0,2,0,3[	15	90%	100%
GRCh38	[0,3,1,0[	0	"N/A"	"N/A"

- **SV Germline**

Tipo de variante	Versão do genoma	Controlos positivos	Verdadeiros positivos	Falsos negativos	Crítérios de aceitação	Sensibilida de
SV da linha germinativa	GRCh37	2	2	0	100%	100%

- **Inserções de Alu**

Versão do genoma	Controlos positivos	Verdadeiros positivos	Falsos negativos	Crítérios de aceitação	Sensibilida de
GRCh37	1	1	0	100%	100%
GRCh38	1	1	0	100%	100%

- **MNV**

Versão do genoma	Type of variant	Crítérios de aceitação	Precisão do exoma	Genoma de precisão
GRCh37	MNV	> 85%	87,0%	86,9%
GRCh38	MNV	> 85%	93,0%	91,5%

- **SNV / Indel Somatic**

- **Sensibilidade**

Versão do genoma	Tipo de variante	Controlos positivos	Verdadeiros positivos	Falsos negativos	Crítérios de aceitação*	Sensibilida de
GRCh37	SNV ($\geq 5\%$)	121	121	0	100%	100%
	Indel ($\geq 5\%$)	21	21	0	100%	100%
GRCh38	SNV ($\geq 5\%$)	121	121	0	100%	100%
	Indel ($\geq 5\%$)	21	21	0	100%	100%

*O critério de aceitação para a deteção de variantes é definido como 100% de sensibilidade para SNVs e indels com um VAF esperado superior a 5%.

○ Precisão

			Precisão			
			GRCh37		GRCh38	
Gene	Variant	Frequência esperada	Média *	Viés	Média frequência *	Desvio
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%

* A média das frequências é obtida a partir de 12 réplicas da amostra HD200.

● SNV/Indel Somatic LF

Genoma versão	Controlo controles	Verdadeiros falsos	Falsos negativos	Crítérios de aceitação *	Sensibilida de
GRCh37	42	42	0	100%	100%
GRCh38	42	42	0	100%	100%

*O critério de aceitação para a deteção de variantes é definido como uma sensibilidade de 100% para SNVs e indels com um VAF esperado superior a 0,1%.

● Somatic Large Panels de CNV

Versão do genoma	Tipo de variante	Controlos positivos	Verdadeiros positivos	Falsos negativos	Crítérios de aceitação	Sensibilida de
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 metilação

Versão do genoma	CrITÉrios de aceitação	Concordância ponderada
GRCh37	>99%	99,88%

- Fusão de ADN somático

Genoma versão	Controlo controles	Verdadeiros falsos	Falsos negativos	Aceitação CrITÉrios	Sensibilida de
GRCh37	9	9	0	100%	100%
GRCh38	9	9	0	100%	100%

- Fusão de ADN somático LBx

Versão do genoma	Fração tumoral	Controlos positivos	Verdadeiros positivos	Falsos negativos	CrITÉrios de aceitação	Sensibilida de
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

Versão do genoma	CrITÉrios de aceitação	Sensibilidade	Precisão
GRCh37	>80%	83%	85%

- MSI

Versão do genoma	Controlos positivos	Controlos negativos	Verdadeiros positivos	Falsos negativos	CrITÉrios de aceitação	Sensibilida de	Precisão
GRCh37	22	24	22	0	100%	100%	100%
GRCh38	6	17	6	0	100%	100%	100%

- GiP

Genoma versão	Controlo controles	Verdadeiros falsos	Falsos negativos	Aceitação CrITÉrios	Sensibilida de
GRCh37	171	165	6	90%	96,5%

Genoma versão	Controlo controles	Verdadeiros falsos	Falsos positivos	Aceitação CrITÉrios	Precisão
GRCh37	171	165	7	90%	95,9%

- **Sentieon SNV / Indel Germline**

Versão do genoma	Tipo de variante	Critérios de aceitação*	Sensibilidade Exome	Sensibilidade 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 a deteção de variantes localizadas nas regiões de alta confiança de NA12878 e NA24385

Versão do genoma	Tipo de variante	Critérios de aceitação*	Precisão Exome	Precisão do 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 a deteção de variantes localizadas nas regiões de alta confiança de NA12878 e NA24385.

- **CNV WGS**

Versão do genoma	Controlos positivos	Verdadeiros positivos	Falsos positivos	Critérios de aceitação	Sensibilidade
GRCh38	Duplicação	1 kb - 5 kb	8	10%	37,5%
	Duplicação	≥ 5 kb	3	10%	100%
	Deleção	1 kb - 5 kb	321	80%	61,7%
	Deleção	≥ 5 kb	132	80%	92,4%

Embora a sensibilidade para pequenas deleções (1-5 kb) fique abaixo dos critérios de aceitação, esta redução constitui um compromisso aceitável para maximizar a deteção de eventos de grande impacto e duplicações.

Versão do genoma	Controlos positivos	Verdadeiros positivos	Falsos positivos	Critérios de aceitação	Precisão
GRCh38	Duplicação	1 kb - 5 kb	8	"NA"	30%
	Duplicação	≥ 5 kb	3	"NA"	75%
	Deleção	1 kb - 5 kb	321	"NA"	94,3%
	Deleção	≥ 5 kb	132	"NA"	92,4%

- **Sentieon SNV/Indel Somatic**

Versão do genoma	Tipo de variante	Controlos positivos	Verdadeiros positivos	Falsos positivos	Critérios de aceitação	Sensibilidade
GRCh38	SNV	39	37	2	85%	95%
	Indel	21	16	5	75%	76%

- **RNA de fusão**

Genoma Versão	Controlo controles	Verdadeiros falsos	Falsos negativos	Aceitação Critérios	Sensibilidade
GRCh37	20	20	0	100%	100%
GRCh38	20	19	1	95%	95%

- **RNA de splicing**

Genoma Versão	Positivo controles	Verdadeiros falsos	Falsos negativos	Aceitação Critérios	Sensibilidade
GRCh37	3	3	0	100%	100%
GRCh38	3	3	0	100%	100%

- **Panel reduzido de CNV somáticas**

Genoma Versão	Positivo controles	Verdadeiros falsos	Falsos negativos	Aceitação Critérios	Sensibilidade
GRCh37	4	4	0	100%	100%
GRCh38	4	4	0	100%	100%

- **Estatuto IGHV IMGT**

Genoma versão	Positivo Controles	Verdadeiros falsos	Falsos negativos	Aceitação Critérios	Sensibilidade
GRCh37	122	116	6	95%	95,9%

- **Estado do IGHV igblast**

Versão do genoma	Controlos positivos	Verdadeiros positivos	Falsos negativos	Critérios de aceitação	Sensibilidade
GRCh37	122	117	5	95%	95,1%

- **Subconjuntos de IGHV**

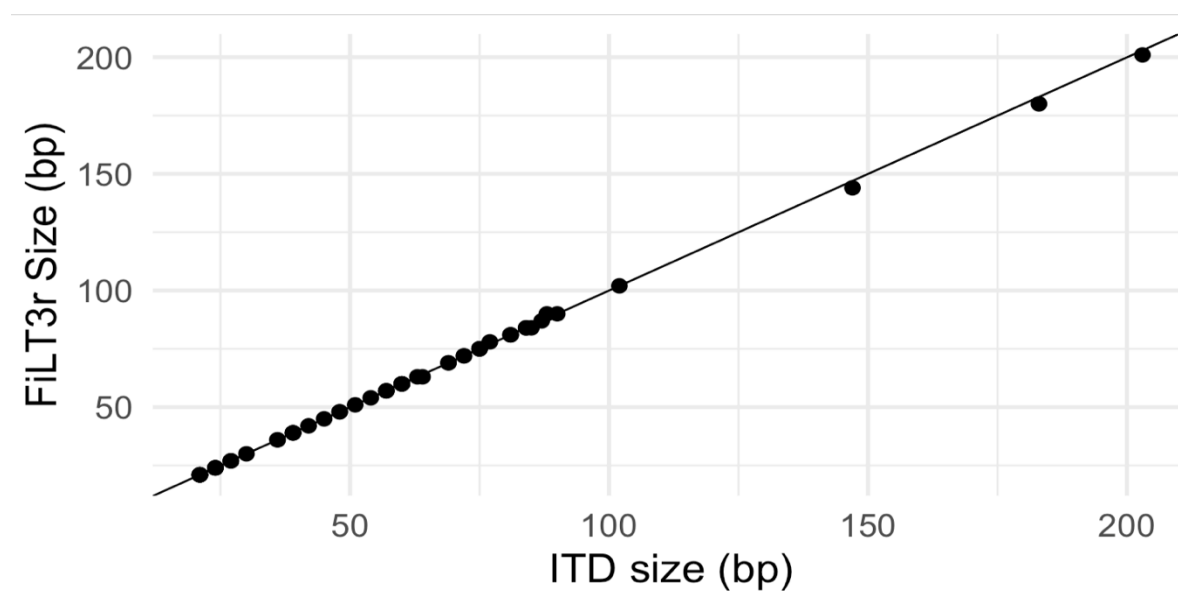
Genoma versão	Controlo controles	Verdadeiros falsos	Falsos negativos	Aceitação Critérios	Sensibilidade
GRCh37	61	10	0	95%	100%

Genoma versão	Controlo controles	Verdadeiros falsos	Falsos negativos	Aceitação Critérios	Precisão
GRCh37	61	10	0	95%	100%

- CFH

Genoma versão	Controlo controles	Verdadeiros falsos	Falsos negativos	Aceitação Critérios	Sensibilidade
GRCh38	4	4	0	100%	100%

- FLT3



Tamanho do evento FLT3-ITD observado vs. teórico. **Coefficiente de Pearson $R^2 > 0,99$**

Os resultados mostram 100% de sensibilidade na deteção de eventos FLT3-ITD e uma correlação quase perfeita entre os tamanhos de evento esperados e observados a partir da análise.

- **UPD**

- **Sensibilidade**

Tipo de análise	Versão do genoma	Controlos positivos	Controlos negativos	Verdadeiros positivos	Falsos negativos	Critérios de aceitação	Sensibilidade
Singleton	GRCh37/38	7	23	7	0	80%	100%
Trio	GRCh37/38	7	2	6	1	80%	86%

- **Precisão**

Tipo de análise	Versão do genoma	Controlos positivos	Controlos negativos	Verdadeiros positivos	Falsos positivos	Critérios de aceitação	Precisão
Singleton	GRCh37/38	7	23	7	0	80%	100%
Trio	GRCh37/38	7	2	6	0	80%	100%

4.6.1.2. Módulos de Anotação

- **Pontuação DiagAI**

Conjunto de dados	Análise de tipo	Métrico avaliado	Aceitação Critérios	Métrica valor
Exomas positivos para doenças raras	"Germline"	Classificação mediana	= 1	1
Exomas positivos para doenças raras	"Germline"	top_10	> 90%	95%
Exomas de CNV de doenças raras	"Germline"	Classificação mediana	= 1	1
Exomas com CNV positivos para doenças raras	Linha germinativa	top_1	> 80%	83,3%

- **Pontuação somática DiagAI**

Conjunto de dados	Tipo de análise	Métrica avaliada	Critérios de aceitação	Valor da métrica
Câncer positivo TSO500	"Somatic"	top_10	> 90%	100%

- **Listas de pré-seleção do DiagAI**

Conjunto de dados	Análise de tipo	Métrica avaliada	Aceitação Critérios	Valor da métrica
Exomas positivos para doenças raras	Linha germinativa	Sensibilidade	> 90	95%
Exomes positivos para doenças raras	Linha germinativa	Comprimento médio	< 25	18,475
Exomes positivos para doenças raras	Linha germinativa	Comprimento mediano	< 25	17

- **DiagAI smartpicks**

Conjunto de dados	Análise de tipo	Métrica avaliada	Aceitação Critérios	Valor da métrica
Exomas positivos para doenças raras	"Germline"	Sensibilidade	> 40%	45%
Exomas positivos para doenças raras	"Germline"	precisão	> 80%	78%

- **ACMG CNV**

Conjunto de dados	Tipo de análise	Métrica avaliada	Critérios de aceitação	Valor da métrica
Conjunto de dados de CNV anotado por especialistas da ACMG	"Germline"	Sensibilidade	> 70%	76%

- **SNV da ACMG**

Conjunto de dados	Tipo de análise	Métrica avaliada	Critérios de aceitação	Valor da métrica
Conjunto de dados anotado por especialistas da ACMG	'Germline'	precisão	> 50%	56%

- **DiagAI UP2**

Conjunto de dados	Análise de tipos	Métrica avaliada	Critérios de aceitação	Métrica valor
Variantes causadoras de doenças	Linha germinativa	Sensibilidade	90%	92%

- **DiagAI UP2 SV**

Conjunto de dados	Tipo de análise	Métrica avaliada	Critérios de aceitação	Métrica valor
Variantes causadoras de doenças	Linha germinativa	Sensibilidade	90%	98%

- **DiagAI Phenogenius**

Conjunto de dados	Type of analysis	Métrica avaliada	CrITÉRIOS de aceitação	Métrica valor
Associações de sintomas genéticos	"Germline"	top_100	50%	49,8%
Associações entre genes e sintomas	Linha germinativa	top_200	60%	60,7%
Associações entre genes e sintomas	"Germline"	top_500	75%	77,0%
Associações entre genes e sintomas	"Germline"	Mediana no top 100	<20	15
Associações entre genes e sintomas	"Germline"	Média no top 100	<30	25

- **Compermed**

Conjunto de dados	Tipo de análise	Métrica avaliado	CrITÉRIOS de aceitação	Métrica valor
Câncer positivo TSO500	Sensibilidade	Sensibilidade	90%	91%

4.6.2. Desempenho clínico

A SeqOne Platform não se destina a ser utilizada para a avaliação de uma condição clínica ou estado fisiológico específico. Embora forneça informações sobre o estado fisiológico de um doente, tais como a identificação de determinadas variantes, não se destina a avaliar condições clínicas ou estados fisiológicos específicos. Por conseguinte, a SeqOne Platform, por si só, não requer uma demonstração de desempenho clínico.

5. Precauções

5.1. Utilização conforme

(Ver secção **Utilizadores a quem se destina**)

É proibida qualquer utilização inadequada da plataforma.

Os utilizadores são responsáveis por verificar se a plataforma está operacional em todos os momentos e por garantir as condições necessárias para o seu bom funcionamento.

5.2. Avisos



Os resultados fornecidos não devem constituir a única base para a alteração dos cuidados prestados ao doente. Consequentemente, os utilizadores devem exercer a devida cautela ao utilizar o software e os seus resultados. Os diagnósticos devem ser confirmados através de vários exames clínicos e médicos.

A investigação de doenças por NGS pode revelar achados secundários ou incidentais, tais como predisposição para o cancro, a partir de genes não relacionados com a condição primária (Kuhlen e Borkhardt 2015). Apesar das declarações de política de agências como a AMP e a ACMG (Green et al. 2013; Allyse e Michie 2013; Conselho de Administração da ACMG 2015; Hegde et al. 2015), as diretrizes nacionais sobre a comunicação de achados incidentais variam. Por conseguinte, é crucial obter um documento de «consentimento informado» (Anderson et al. 2016). O médico responsável pelo tratamento é responsável por garantir que o paciente ou o seu representante legal compreenda os potenciais resultados e riscos antes de solicitar qualquer teste de NGS.

Este produto não garante a deteção de todas as variantes genéticas presentes numa amostra. A probabilidade de detetar uma variante de interesse depende, por natureza, da qualidade e das características da amostra de entrada (por exemplo, integridade da amostra, preparação da biblioteca, profundidade e uniformidade do sequenciamento, filtragem bioinformática e tipo ou contexto da variante). Um resultado negativo não exclui definitivamente a presença de uma variante patogénica se a sua frequência (VAF) for inferior aos limiares de deteção do módulo analítico utilizado. A deteção fiável de rearranjos genómicos em grande escala (variantes estruturais, incluindo CNVs e deleções de cromossomas inteiros) não é garantida. Isto deve-se à limitação inerente aos dados de sequenciação de leitura curta de última geração (NGS), o que torna estes eventos difíceis de identificar. Como as leituras NGS são curtas (aproximadamente 150-200 pb), esta restrição é uma característica da própria tecnologia de sequenciação e não constitui uma limitação da SeqOne Platform. (Ver Tan R. et al., Hum Mutat. Julho de 2014; 35(7):899-907. doi: 10.1002/humu.22537 e referências nele contidas). Os utilizadores devem avaliar a adequação do

conjunto de trabalho selecionado e a qualidade do sequenciamento em relação às variantes esperadas, para garantir que o contexto clínico apropriado foi abrangido.

As decisões clínicas baseadas nos resultados do rastreio de recém-nascidos ou de portadores devem ser tomadas com extrema cautela e nunca se basear exclusivamente nos resultados deste software. No rastreio de recém-nascidos, um resultado positivo requer uma correlação clínica imediata com os sintomas do bebê e testes de confirmação urgentes antes de iniciar tratamentos que alterem a vida ou restrições alimentares. No rastreio de portadores, um resultado negativo não elimina o risco de ser portador, mas reduz-o a um nível de risco residual, uma vez que certas variantes genômicas podem ficar abaixo do limite de detecção do software. Como os resultados deste dispositivo influenciam diretamente escolhas reprodutivas irreversíveis e cuidados neonatais críticos, todos os resultados passíveis de ação devem ser verificados utilizando um método alternativo validado e interpretados por um profissional qualificado no contexto do historial clínico completo do paciente e dos dados bioquímicos ou fenotípicos relevantes.

Variantes recém-descobertas que ainda não estejam exaustivamente documentadas na literatura científica podem ser classificadas incorretamente.

Os resultados são precisos no momento da análise e refletem o conhecimento científico atual. No entanto, a repetição da análise numa versão mais recente da plataforma pode produzir resultados diferentes.

Os resultados de classificação da SeqOne Platform visam alinhar-se com o consenso comum. Tenha em atenção que quaisquer resultados gerados após a alteração das definições de classificação estão fora do âmbito da CE-IVD e, por conseguinte, não são validados.

Este software incorpora componentes de inteligência artificial (IA). Os utilizadores devem estar cientes das seguintes limitações inerentes a estas ferramentas baseadas em IA e tê-las em conta ao interpretar os resultados:

- Os dados de treino são predominantemente derivados de populações ancestrais europeias/caucasianas devido à composição das bases de dados de referência clínicas atuais;
- O desempenho pode variar para variantes raras com representação limitada nos dados de treino;
- Os índices de confiança refletem a incerteza do modelo, não a certeza clínica; devem ser interpretados em conjunto com o contexto clínico;
- O desempenho pode deteriorar-se com a deriva de dados proveniente de novas plataformas de sequenciação ou protocolos de preparação de bibliotecas não representados nos dados de treino;
- O desempenho da priorização de genes do DiagAI Phenogenius é avaliado com base em associações gene-sintoma e depende da qualidade e da exaustividade da anotação de termos HPO pelo médico solicitante.

5.3. Limitações

"Carrier screening" está sujeito a legislação nacional específica e a regulamentos éticos. É da responsabilidade do laboratório e do prestador de cuidados de saúde garantir que todos os requisitos legais — incluindo aconselhamento genético obrigatório, pedido por um médico e consentimento informado — sejam cumpridos na sua jurisdição específica.

5.4. Avarias

Em caso de avaria:

- Pare imediatamente de utilizar o software
- Tente identificar ou eliminar as causas seguindo o procedimento de resolução de problemas descrito neste guia (consulte a secção **Resolução de problemas e mensagens de erro**)
- Se não conseguir identificar ou eliminar a causa utilizando este guia, envie um email para a equipa de apoio através do endereço support@seqone.com.

6. Requisitos mínimos

6.1. Hardware

Os módulos de software da plataforma SeqOne são instalados como parte de um arranjo cliente-servidor. O fornecedor de serviços deve ser um servidor Linux para garantir a compatibilidade com a maioria das ferramentas habitualmente utilizadas em bioinformática. Cada item do sistema é encapsulado num ambiente de tempo de execução, como o contentor Docker®, que pode ser implementado em qualquer plataforma Docker® compatível, independentemente da arquitetura de hardware subjacente. Os utilizadores devem ter um computador com ligação à Internet e um navegador Web (ver abaixo).

Os requisitos mínimos do sistema para um computador cliente são os seguintes:

- 4 GB DE RAM
- Processador dual-core
- Acesso à Internet
- Navegador Chrome (a partir da versão 36.0.1985) ou Firefox (a partir da versão 31) ou Microsoft Edge (todas as versões baseadas no Chromium).

6.2. Rede

A velocidade mínima de carregamento/d Descarregamento recomendada para a utilização da plataforma através da Internet é de 100 Mbps. O protocolo HTTPS é utilizado para a ligação à plataforma. Para aceder corretamente à plataforma, recomendamos que peça ao seu serviço informático que autorize os seguintes endereços :

*.eu.seqone.com

*.common.seqone.com

7. Autenticação

7.1. Iniciar sessão

Por razões de segurança, os utilizadores só podem aceder à plataforma depois de iniciarem sessão no ecrã inicial.

O ecrã inicial está disponível através do seguinte URL:

- Para a Europa e África: <https://auth.eu.seqone.com/>
- Para a Suíça: <https://auth.ch.seqone.com/>
- Para a Ásia e a Oceânia: <https://auth.au.seqone.com/>
- Para as Américas: <https://auth.us.seqone.com/>

Ao iniciar sessão no ecrã inicial, tem de introduzir o seu nome:

- Login ou endereço de correio eletrónico no campo "Username" ou "email"
- Palavra-passe no campo "Password".

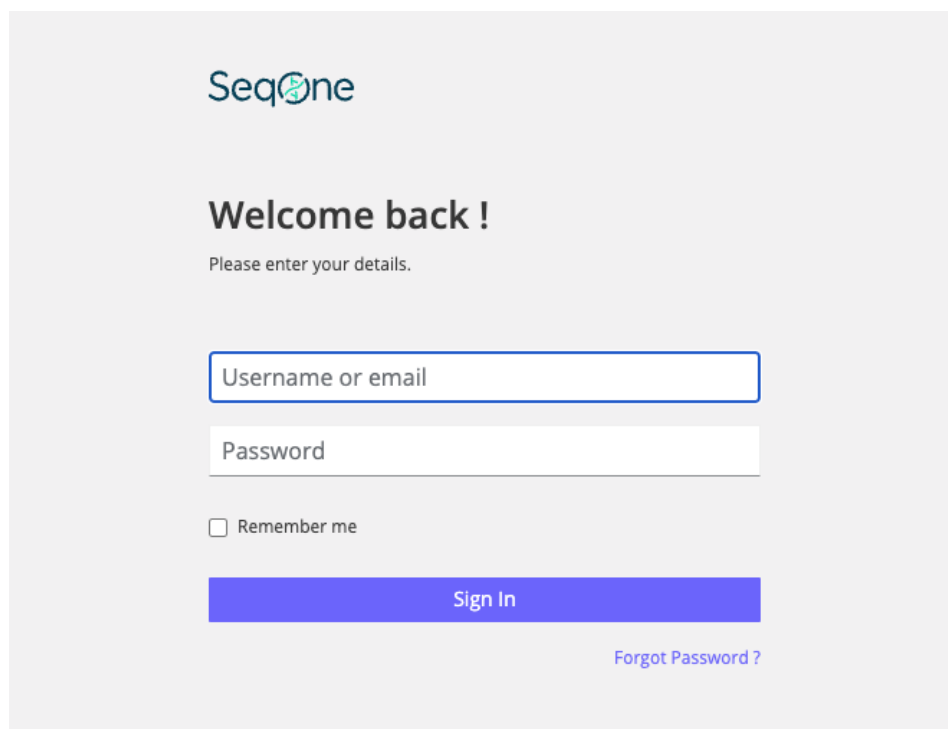
The image shows a login page for SeqOne. At the top is the SeqOne logo. Below it, the text "Welcome back !" is displayed in a large, bold font. Underneath, in a smaller font, it says "Please enter your details." There are two input fields: the first is labeled "Username or email" and the second is labeled "Password". Below these fields is a checkbox labeled "Remember me". At the bottom of the form is a blue button labeled "Sign In". To the right of the button, there is a link that says "Forgot Password ?".

Figura 1: Página de início de sessão

Se a autenticação de dois factores estiver activada para a sua conta, terá de introduzir um código de verificação de 6 dígitos enviado por email para o endereço associado à sua conta SeqOne. Este código é válido durante 30 minutos.

Se tiver acesso a vários espaços de trabalho ou entidades, ser-lhe-á pedido que selecione a entidade à qual se pretende ligar através de um menu pendente em "Choose a workspace".

Pode alterar a sua password clicando em "Forgot Password". Siga as instruções introduzindo o endereço de e-mail associado à sua conta SeqOne e receberá um e-mail que lhe permitirá repor a sua palavra-passe.

A política de palavras-passe segue as seguintes regras:

- Comprimento mínimo: 12 caracteres.
- Complexidade: as palavras-passe devem incluir pelo menos uma letra maiúscula, uma letra minúscula, um número e um carácter especial. Quanto maior for o conjunto de caracteres, mais forte será a palavra-passe.
- Histórico de palavras-passe: não reutilizar as palavras-passe anteriores.
- Sem comprimento máximo: para permitir a utilização de frases secretas.

Não é possível aos utilizadores alterarem o seu próprio ID de utilizador.

7.2. Terminar sessão

Os utilizadores podem terminar a sessão clicando no logótipo da conta de utilizador no canto superior direito da página e depois em "Logout".

7.3. Página inicial

Após a autenticação, o utilizador acede à página do painel de controlo, que está dividida em duas partes:

- uma faixa azul de pesquisa e ação, acessível em toda a navegação, da esquerda para a direita :
 - o ícone SeqOne para regressar ao painel de controlo em qualquer altura,
 - uma barra de pesquisa para acesso direto a uma determinada análise,
- um ícone de pesquisa VKB  para procurar variantes avaliadas na base de dados VKB (ver secção **Avaliação de variantes**),
- um ícone  que abre a "To do list" do utilizador, a partir da qual este pode aceder às análises que lhe foram atribuídas (ver secção **Atribuição de análises para interpretação**),
- uma barra de menu, disponível através do ícone  , que dá acesso aos parâmetros do utilizador, ao painel do administrador, ao manual do utilizador, aos termos de utilização, ao endereço de email de apoio e que permite ao utilizador terminar a sessão (ver secção **Gestão de contas de utilizador**).

- o ecrã principal para gerir projetos e análises.

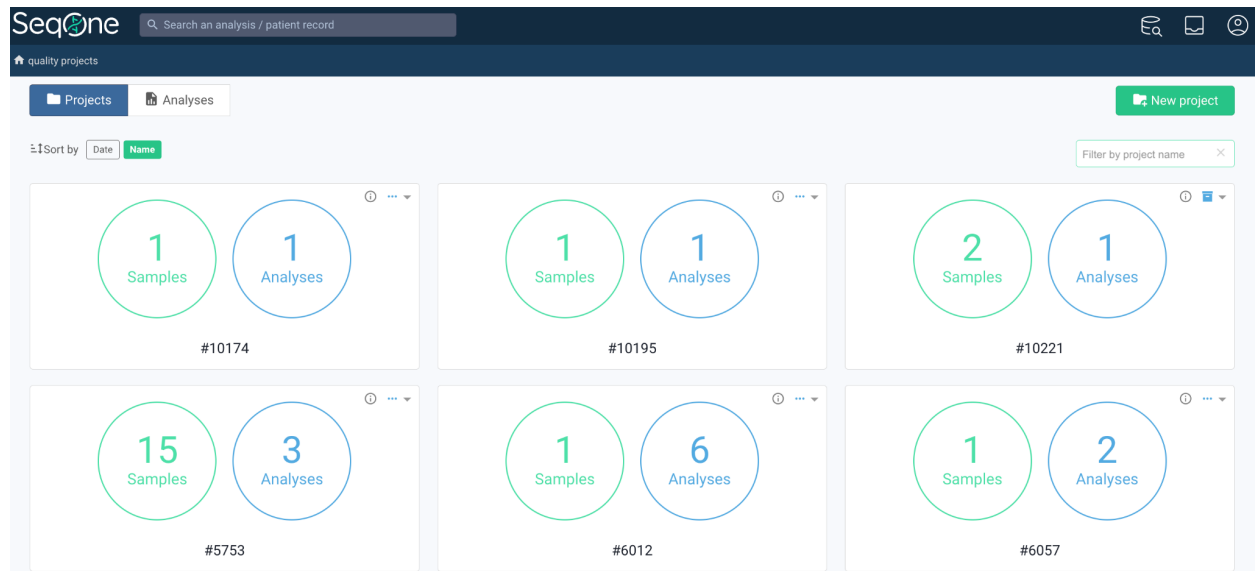


Figura 2: Painel de controlo do utilizador

Pode aceder ao painel de controlo do utilizador em qualquer altura, clicando no logótipo SeqOne.

7.4. Gestão de contas de utilizador

Pode gerir a sua conta de utilizador clicando no logótipo no canto superior direito da barra de tarefas.

Nesta secção, o utilizador tem a possibilidade de:

- aceder às configurações da conta clicando em "My account"
- aceder aos parâmetros do utilizador clicando em "Entity Settings"
- alterar a entidade à qual o utilizador está ligado clicando em "Entity" e seleccionando a entidade pretendida em "Choose your workspace",
- aceder ao painel de administrador clicando em "Admin", permitindo o acesso à lista de utilizadores com acesso à entidade e a criação, modificação ou eliminação de contas de utilizador.
- clicar em "About" para aceder ao manual do utilizador em vários idiomas, informações sobre a versão do software SeqOne e detalhes de contacto,
- aceder aos termos da sua licença SeqOne clicando em "Terms of Use",
- clicar em "Support" para aceder ao endereço de e-mail de suporte para quaisquer questões ou problemas encontrados na plataforma,

- aceder ao atalho clicando em "Keyboard Shortcuts"
- alternar entre a versão padrão e a versão de visualização da aplicação clicando em "SeqOne Mode" e seleccionando "SeqOne" (standard mode) ou "SeqOne Preview" (preview mode). O modo atualmente ativo é indicado por uma marca de verificação.
- terminar sessão clicando em "Logout".

7.4.1. Gestão de "My Account"

Na secção "My account", o utilizador pode:

- carregar a sua assinatura para a incluir no relatório clínico,
- alterar a sua palavra-passe clicando em "Change your password" (abre um novo separador) na secção "Manage your credentials".

7.4.2. Gerir as definições do utilizador - definições

Nas configurações de usuário, o utilizador pode gerir as definições, arquivos e configurações descritos abaixo.

Entity

A entidade refere-se à instituição ou equipa que adquiriu uma licença da plataforma SeqOne. A plataforma foi concebida para ser utilizada por vários membros do mesmo grupo. A partir das "Settings" do utilizador, o separador "Account" é utilizado para gerir as informações relativas à entidade.

Users

Para este separador, consulte a **Gestão administrativa - apresentação da interface**.

Worksets

Este separador lista os diferentes worksets de análise disponíveis na entidade à qual o utilizador está ligado. Uma vez que podem coexistir várias versões de determinados pipelines na plataforma SeqOne, o utilizador pode definir a partir deste menu qual a versão que deve ser pré-seleccionada ao iniciar a análise.

As modificações efectuadas nesta lista são controladas, de modo a que o utilizador que definiu a versão a lançar por defeito de um workset seja automaticamente apresentado, bem como a data da modificação.

Manifest

O manifesto apresenta as regiões genómicas de interesse (ROI) como um ficheiro BED, que pode ser acompanhado pelas respectivas 'Annotations' associadas. Pode carregar um novo manifesto através do separador *Manifesto* nas definições do utilizador, clicando em "Add manifest" e seleccionando o ficheiro de manifesto que pretende carregar. Uma vez carregado,

este manifesto estará disponível para todos os utilizadores quando criarem um novo projeto e quando aplicarem um 'in silico' panel numa análise.

Panels - Transcript

Os utilizadores podem carregar uma nova lista de transcrições de referência (RefSeq) através do separador *Transcript*.

Panels - Genes

Este separador permite-lhe importar ou gerar painéis de genes com base no NCBI GeneID, clicando em "Add gene list".

O formato de ficheiro esperado é um ficheiro tabular (por exemplo, um ficheiro .tsv) sem cabeçalhos de coluna, o que é detalhado na plataforma. A primeira coluna do ficheiro esperado é dedicada aos identificadores de genes (NCBI Entrez gene ID), a segunda coluna (opcional) contém o transcript de interesse (RefSeq ID) e a terceira coluna (opcional) contém o símbolo HGNC do gene. Este ficheiro também pode ser gerado inserindo a lista desejada de genes, separados por vírgula, no campo 'Gene symbols'.

Marker

Este separador é utilizado para importar uma lista de marcadores de identidade genética clicando em 'Add marker config'. O formato de ficheiro esperado é um ficheiro tabular (.tsv ou .txt) com um cabeçalho de coluna contendo 5 colunas com os seguintes detalhes: cromossoma (CHR), identificador de variante ou rsID (ID), posição genómica (POS), alelo de referência (REF) e alelo alternativo (ALT). O ficheiro não deve conter mais de 100 linhas, incluindo o cabeçalho.

Protocols

Um protocolo de análise é uma combinação predefinida de tarefas concebida para padronizar a interpretação das análises. O separador correspondente permite aos utilizadores visualizar a lista de protocolos disponíveis na entidade, as suas versões, e criá-los.

Para criar um protocolo, clique em "New protocol", depois adicione e ordene as tarefas que correspondam aos SOP do seu laboratório. Para editar um protocolo existente, clique no ícone do lápis ao lado do mesmo e guarde a nova versão do protocolo.

Selecione um protocolo no lançamento da análise, ou anexe-o mais tarde a partir do separador Workflow no painel de análise. Marque as tarefas como concluídas diretamente no painel de análise. Cada tarefa concluída regista automaticamente o utilizador e a data para uma rastreabilidade total.

Reporting - Global settings

A secção Global settings do separador Reporting permite aos utilizadores ativar ou desativar o bloqueio automático da análise quando o estado muda para "Reporting Validating", e personalizar as secções de classificação e relatório. Para o fazer, em Germline / Somatic, clique em "New section", introduza um nome de secção e selecione uma etiqueta para associar à classificação desejada. As variantes que correspondam a uma das etiquetas ou classificações selecionadas são adicionadas automaticamente a esta secção. Deve primeiro criar etiquetas no separador correspondente para utilizar esta funcionalidade.

Reporting - Template list

O separador Reporting em "Template list" permite aos utilizadores configurar os modelos utilizados na exportação dos resultados da análise. A partir deste separador, pode personalizar a informação incluída nos relatórios gerados.

Tags

O separador "Tags" permite aos utilizadores gerir etiquetas personalizadas para organizar e filtrar variantes na VKB e no relatório. Os utilizadores podem criar etiquetas clicando em "Add tag", ou renomear e eliminar etiquetas que estão disponíveis para todos os utilizadores da entidade utilizando os ícones de caneta e lixo correspondentes.

API Keys

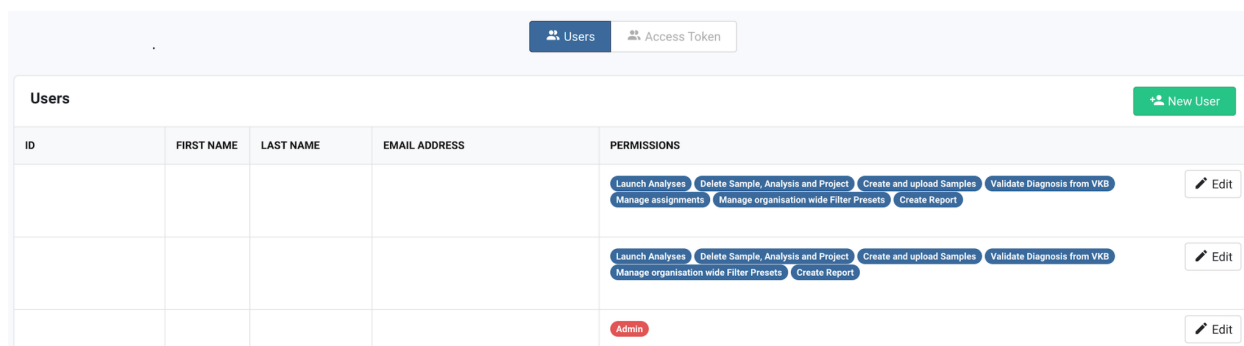
Uma chave API é um token de autenticação exclusivo que permite que aplicações ou scripts externos interajam com a plataforma SeqOne de forma programática. No separador API Keys, pode visualizar todas as chaves existentes, juntamente com as informações associadas.

Crie uma nova chave clicando em "+ New API Key". As chaves expiradas podem ser mostradas ou ocultadas utilizando o botão de alternância "Show expired key". Cada chave pode ser eliminada individualmente. As chaves API devem ser mantidas confidenciais, pois concedem acesso à plataforma com as respetivas permissões associadas.

7.4.3. Gestão administrativa - apresentação da interface

Esta interface é dedicada aos administradores de contas e permite-lhes :

- aceder à lista de utilizadores da entidade,
- criar novos utilizadores,
- modificar os direitos do utilizador: criar projetos e carregar amostras, lançar análises, apagar amostras, análises ou projetos, avaliar variantes no VKB, criar relatórios, filtrar a gestão de perfis, gerir atribuições.
- ativar a dupla autenticação,
- desativar uma conta de utilizador.



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: Painel do administrador

O ID corresponde ao nome de utilizador do utilizador. Deve estar em minúsculas e não conter acentos nem espaços; são aceites caracteres especiais como o ponto e o sublinhado. É aconselhável seguir a nomenclatura primeiro nome.último nome quando se cria um novo utilizador.

No que diz respeito ao endereço de correio eletrónico do utilizador, desaconselhamos vivamente a utilização de um endereço de correio eletrónico genérico por razões de segurança.

Como as permissões estão desactivadas por defeito, cabe ao administrador activá-las 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: Permissões do utilizador

Também pode aumentar a segurança activando a autenticação dupla. Esta opção envia um código único ao utilizador sempre que este se liga à plataforma.

A função "Manage organization wide Filter Presets" permite ao utilizador modificar perfis de filtro públicos. Se o utilizador não tiver esta função, só pode modificar perfis de filtro privados.

As permissões são geridas por um administrador de entidade a partir do painel Admin. Quando um utilizador é criado, todas as permissões são desativadas por defeito, e o administrador deve ativar manualmente os direitos necessários para esse utilizador. Os direitos atribuídos a um utilizador aplicam-se a todas as entidades (espaços de trabalho) às quais o utilizador tem acesso na plataforma SeqOne (por exemplo, se o utilizador puder aceder a várias entidades, o mesmo conjunto de permissões aplica-se a todas elas).

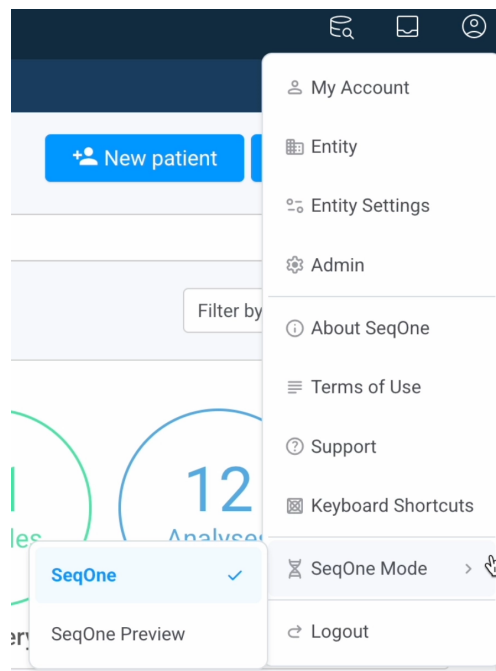
Se receber uma mensagem de erro, contacte o apoio ao cliente através do endereço support@seqone.com.

7.4.4. Alternando entre os modos Main e Preview

Conectar-se ao SeqOne através do URL de login padrão leva os usuários ao modo “Main” de sua entidade. Esse modo padrão hospeda conjuntos de trabalho com marcação CE e outras certificações locais.

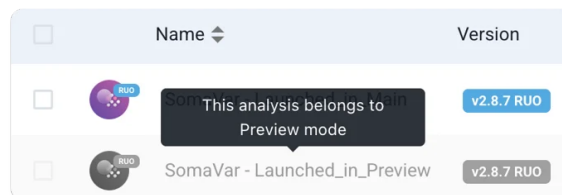
Para acesso antecipado a novos recursos e versões de conjuntos de trabalho, mude para o modo “Preview” (RUO: Research Use Only) indo até a parte inferior do menu de configurações e, em SeqOne One, selecionando “Preview”.

Os modos “Main” e “Preview” compartilham os mesmos dados principais (projetos, amostras, análises), permissões, recursos (arquivos, modelos, manifestos, listas de genes), conhecimento (VKB) e frequências (dados de frequência de amostras, ocorrências).



Ao navegar no modo “Preview”, o logotipo “SeqOne Preview RUO” aparece como um lembrete de que este modo contém recursos e conjuntos de trabalho RUO.

No modo “Main”, as análises iniciadas em “Preview” aparecem esmaecidas. Se você clicar em uma dessas análises esmaecidas, a plataforma solicitará que você mude para o modo Preview para visualizá-la.



8. Descrição das etapas operacionais

Etapas	Descrição
1	Criação de projetos As análises realizadas na plataforma SeqOne estão organizadas em projetos, que definem as características das amostras analisadas através de metadados introduzidos pelo utilizador (tipo de sequenciação, método de preparação, genes de interesse).
2	Upload de amostras As amostras são carregadas como ficheiros de sequenciação em bruto (FASTQ), que devem ser colocados no projeto designado. Uma vez carregados os ficheiros, uma etapa de verificação garante a integridade dos ficheiros enviados antes de qualquer análise.
3	Lançamento da análise Uma vez carregados os dados nos servidores SeqOne, o utilizador pode escolher entre uma seleção de aplicações ou "worksets", em função do tipo de amostra analisada e dos tipos de alterações genómicas procuradas (SNV, CNV, fusões de genes, variantes de splicing, etc.). Estas aplicações efectuem análises bioinformáticas.
4	Controlo da qualidade dos dados As aplicações disponíveis na plataforma SeqOne incluem vários módulos dedicados à análise da qualidade dos dados, utilizando várias métricas de sequenciação, alinhamento e profundidade de cobertura.
5	Interpretação dos resultados da análise As variações genómicas detetadas na amostra são apresentadas em vários separadores ou "viewers", onde o utilizador pode procurar a variante de diagnóstico responsável pela patologia do paciente.
6	Obtenção de recomendações terapêuticas e ensaios clínicos associados à patologia do paciente.
7	Emissão de um relatório Os resultados da análise podem ser exportados como um relatório PDF", que pode depois ser descarregado.

9. Criação de projetos

As análises realizadas na plataforma SeqOne são organizadas em projetos, que definem as características das amostras analisadas através de metadados introduzidos pelo utilizador (tipo de sequenciação, método de preparação, genes de interesse).

9.1. Criar um projeto

Depois de clicar em "New project", é pedido aos utilizadores que introduzam os valores de vários tipos de informação, ou metadados, que são utilizados para descrever o projeto e as amostras associadas.

- **Descrição do projeto**
- **Sequenciação:**
 - Assay type: tipo de análise (Germline ou Somatic),
 - Os projetos Germline (*Assay type Germline*) apresentam uma caixa de verificação UMI Trimming (Corte de UMI), permitindo que as amostras que contêm UMIs sejam processadas. Ao selecionar esta opção, as UMIs serão cortadas das sequências FASTQ, tal como no modo UMI disabled para worksets somáticos (consulte a secção **Escolher o modo de processamento UMI**).
 - Sequencing platform: a plataforma de sequenciação
 - Library type: tipo de sequenciação de read efectuada (*Paired-end* ou *Single-end*),
 - Data type: DNA, RNA, ctDNA ou Methylation Sequencing,
 - Se for selecionado ADN para o tipo de dados, o método de sequenciação deve ser especificado utilizando o menu pendente "*DNA Sequencing method*", bem como o método de amplificação (Capture ou Amplicon) através de "*Amplification method*".
 - File input type: *Ficheiros Fastq*, *ficheiros VCF*
 - Genome: a versão do genoma de referência (GRCh37 ou GRCh38),
 - Low-Quality Mapping Region: a chamada de variantes em regiões de alinhamento de baixa qualidade, considerando as 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: o manifesto no caso de um painel de genes,
 - Marcador de ID (opcional): o ficheiro do marcador de identidade.
- **Amostras:**

- (Opcional) Podem ser adicionados novos metadados de amostra clicando em "*Também pode criar os seus próprios metadados.*" Quando solicitado, introduza o nome e o valor dos metadados.

Uma vez introduzidos os diferentes metadados do projeto, os utilizadores são convidados a criar o seu projeto, clicando no botão "*Create project*".

9.2. Gestão de projetos

Através do painel de controlo, os utilizadores podem gerir os diferentes projetos que foram criados na plataforma. Em particular, podem:

- Aceder aos projetos através do separador "*Projects*", bem como às amostras e análises de cada projeto, clicando nos logótipos "*Samples*" e "*Analysis*", respetivamente.
- Aceder diretamente às análises através do separador "*Analysis*".
- Criar um novo projeto clicando em "*New project*" (ver a secção **Criar um projeto**).

9.3. Eliminar um projeto

Para eliminar um projeto, os utilizadores devem:

- Selecionar o projeto
- Clique no separador "*Settings*" e depois em "*Delete project*"
- Confirmar assinalando "Yes" na mensagem apresentada. Esta ação é irreversível.

9.4. Dados de ADN (sequenciação de ADN, sequenciação de ADNct)

Low-Quality Mapping Region. Ao sequenciar painéis de genes, exomas ou genomas, o utilizador pode incluir regiões de alinhamento de baixa qualidade ao chamar variantes. Estas regiões correspondem às chamadas 'MAPQ0' reads, cuja qualidade de alinhamento é igual a 0, tais como pseudogénios e regiões homólogas do genoma.

A opção é aplicável a genes ou regiões-alvo específicos. Para utilizar esta opção, o utilizador é convidado a carregar o manifesto que visa as regiões desejadas ou a seleccioná-lo através de um menu pendente.

Esta opção pode reduzir a qualidade e a exatidão dos resultados. Pode introduzir ruído de fundo na frequência alélica e ter um impacto negativo na priorização de variantes. Pode também aumentar o tempo de análise devido a uma carga computacional mais elevada.

Exome. No caso da sequenciação do exoma, os utilizadores não precisam de especificar o manifesto a utilizar; são convidados a selecionar o fabricante do kit utilizado clicando num dos ícones da categoria Exome kit provider (Fornecedor de kits Exome). Um menu pendente permite

então ao utilizador seleccionar o manifesto pretendido. Se o manifesto pretendido não estiver disponível na plataforma, o utilizador é convidado a contactar o serviço de apoio ao utilizador (support@seqone.com).

Gene panel. No caso de um painel de genes, é pedido aos utilizadores que seleccionem o *método de amplificação*, bem como o manifesto (em formato BED) correspondente aos genes de interesse num menu pendente. Se o manifesto necessário não estiver disponível na plataforma, pode ser carregado seleccionando "Upload new manifest" (*Carregar novo manifesto*) e seguindo as instruções no ecrã.

Shallow WGS, CGH MicroArrays, Whole genome sequencing. No caso da sequenciação do genoma, o utilizador não precisa de especificar qual o manifesto a utilizar.

Capture. Ao seleccionar o método de amplificação *Capture*, está disponível um ícone UMI (*Unique Molecular Identifiers - Identificadores Moleculares Únicos*). Os identificadores moleculares únicos são índices que são adicionados antes da amplificação para ajudar a eliminar duplicados de PCR e reduzir o ruído de fundo após a análise.

Amplicon. Deve ser dada especial atenção aos primers de amplificação aquando da análise de amplicons. Ao criar um projeto para dados de amplicon, os utilizadores são aconselhados a seleccionar um manifesto em que as regiões correspondentes aos alvos não incluam as sequências de primers. Se tal BED não estiver disponível, pode ser seleccionada uma opção para *recortar* um comprimento fixo no final das sequências.

Identity monitoring. Para os métodos de sequenciação correspondentes a painel de genes, exoma, perfil genómico (CGP) ou genoma (shallow WGS) a partir de ficheiros FASTQ, o utilizador é convidado a seleccionar o ficheiro de marcadores de monitorização da identidade a partir de um menu pendente. Se o ficheiro pretendido não estiver disponível na plataforma, pode ser importado seleccionando 'Upload new markers' e seguindo as instruções apresentadas no ecrã (ver também o parágrafo **Marker** na secção **Gerir as definições do utilizador**).

9.5. Dados de ARN (Sequenciação de ARN)

Se RNA for selecionado para o tipo de dados, os utilizadores devem especificar se o seu protocolo retém a informação da cadeia original através do menu pendente *Stranded protocol*'. Se o protocolo for específico da cadeia, a ordem de síntese das cadeias (*Forward-reverse*' ou *reverse-forward*') tem de ser especificada através da opção *Read orientation*' (*Orientação de leitura*). Por último, deve ser especificado o método de sequenciação (painel de genes).

Read orientation. Estão disponíveis duas opções para especificar a orientação da sequência:

- **Forward-Reverse:** o primeiro read (R1) de cada par corresponde à sequência do ARN mensageiro transcrito (strand sense).
- **Reverse-Forward:** o primeiro read (R1) de cada par corresponde à sequência complementar do ARN mensageiro transcrito (cadeia anti-sentido).

Em geral, a orientação das reads é a seguinte para os kits mencionados abaixo:

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

Em caso de dúvida, contactar o serviço de apoio (support@seqone.com).

Uma vez especificado o tipo de dados e configurado corretamente o projeto, os utilizadores são convidados a validar, clicando no botão "Next step".

9.6. Definições do projeto

No separador "Settings" do projeto, é possível :

- modificar o nome e a descrição do projeto nos campos "Project name" e "Project description",
- modificar o perfil de controlo de qualidade (QC),
- alterar o perfil de filtro predefinido na tabela de variantes,
- apagar permanentemente o projeto.

9.6.1. Configurar métricas QC padrão do projeto

O utilizador pode optar por utilizar as métricas QC predefinidas ou definir valores limite personalizados a partir do separador "Settings" do projeto. O utilizador é convidado a contactar o apoio ao utilizador (support@seqone.com) para definir os valores desejados.

Serão aplicadas métricas QC" personalizadas :

- QC checks" na vista "Analyses" do projeto,
- Verificações QC na secção QC auto-validation" do separador QC da análise,
- para o separador "GeneCov" da análise.

9.6.2. Definir o perfil de filtro predefinido do projeto

O utilizador pode optar por definir o perfil de filtro predefinido aplicado à tabela de variantes predefinida a partir do separador Settings do projeto (ver também a secção **Utilizar um perfil de filtro predefinido**).

Pode ser seleccionado um perfil de filtro para cada workset através de um menu pendente, desde que seja público. O perfil de filtro será então aplicado por defeito ao visualizar o quadro de variantes, para todas as análises de projetos, para todos os utilizadores da organização.

Se este parâmetro não for modificado pelo utilizador, quando o projeto for criado, o valor por defeito será o perfil definido pelo SeqOne.

Os perfis de filtro predefinidos podem ser bloqueados clicando no botão "Lock Variants Filters".

10. Carregamento de amostras

10.1. Carregamento de amostras

Os utilizadores podem importar as suas amostras através do ícone 'Add samples' em cada projeto. As amostras podem ser importadas para a plataforma SeqOne de duas maneiras:

A partir de um Computador

Os utilizadores podem carregar os seus ficheiros FastQ diretamente a partir de uma pasta no seu PC, clicando no separador "*From your computer*" a partir do ícone "*Add files*". Os utilizadores também podem arrastar e largar os ficheiros do seu PC para a área dedicada na interface.

Via SDS (SeqOne Data Sync)

O SDS é um módulo premium opcional que permite criar automaticamente projetos, transferir dados, iniciar análises e recuperar dados estruturados (TSV) para integração no sistema de gestão de informações laboratoriais (LIMS).

Este módulo compatível com Linux, MacOS e Windows é um executável que permite aos utilizadores autenticarem-se e interagirem com as APIs SeqOne num ambiente predefinido e altamente seguro.

Requisitos mínimos:

- Linux 1.19 mínimo
- MacOS 10.15 com CPU Intel, no mínimo
- MacOS com CPU Apple Silicon (chip M1), no mínimo
- Windows 10 64bits no mínimo
- 4 GB de RAM (8 GB para Windows)
- Processador quad-core
- Acesso à Internet e à plataforma SeqOne (app.eu.seqone.com)
- Acesso ao sistema de armazenamento de dados (diretamente ou através da rede)

Para utilizar esta solução ou para obter mais informações, contactar a nossa equipa de apoio (support@seqone.com).

Estado de carregamento de amostras

A plataforma SeqOne utiliza diferentes estados para permitir aos utilizadores acompanhar o progresso dos seus carregamentos de amostras:

- *Upload pending*: importação em curso,
- *Verifying sample...*: a conformidade da amostra está a ser verificada,
- *Mapping...*: etapa de alinhamento da amostra,
- *Verified sample*: amostra verificada,
- *Upload failed*: a importação falhou.

Capacidade de carregamento

A capacidade de carregamento de ficheiros (especialmente para ficheiros fastq.gz) depende diretamente da ligação à Internet do utilizador. Por conseguinte, a velocidade da ligação à Internet deve ser tida em conta ao carregar amostras. Consulte a tabela seguinte para ter uma ideia mais clara dos tempos estimados de carregamento:

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

Figura 5: Tempo estimado de carregamento em função do tamanho do ficheiro e da velocidade de ligação à Internet.

10.2. Formato do ficheiro de entrada

10.2.1. Ficheiros de análise primária

O formato do ficheiro a carregar para cada amostra depende da plataforma de sequenciação utilizada. Os nomes dos ficheiros devem estar em conformidade com a convenção de nomenclatura de ficheiros de cada sequenciador.

Os ficheiros têm de ser comprimidos, utilizando a extensão .gz (gzip), antes de poderem ser importados para a plataforma. No caso da sequenciação paired-end", são esperados dois ficheiros, correspondentes às reads de ambas as extremidades de cada molécula.

Quando uma amostra tiver sido corretamente detetada pela plataforma no momento da importação, é apresentado o nome completo do ficheiro. É então possível mudar o nome do ficheiro clicando no mesmo. Os ficheiros Paired-end" são automaticamente combinados numa única amostra.

Os seguintes sequenciadores são suportados pela plataforma:

- [Illumina](#) : formato fastq.gz, apenas "Paired-end",
- [Element Biosciences](#) : formato fastq.gz, apenas em pares,

- *MGI* : formato fastq.gz, apenas "Paired-end". Os nomes de ficheiro esperados são: sample_1.fq.gz e sample_2.fq.gz ou flowcellID_L01_sample_1.fq.gz e flowcellID_L01_sample_2.fq.gz

10.2.2. Ficheiros de análise secundários

O formato do ficheiro a carregar para cada amostra depende da plataforma de sequenciação e da aplicação utilizada para a análise secundária. Os ficheiros devem ser comprimidos com a extensão .gz (gzip) antes da importação para a plataforma, salvo indicação em contrário.

O formato de importação dos ficheiros de análises secundárias gerados com as aplicações Illumina, Oxford Nanopore Technologies, Agilent, AmoyDx, Pillar e Thermofisher depende do workset utilizado (ver secção **Worksets de análise disponíveis a partir de ficheiros de análises secundárias**).

Quando a plataforma detecta corretamente uma amostra durante a importação, apresenta o nome completo do ficheiro. É possível mudar o nome do ficheiro clicando no mesmo. Os ficheiros que partilham o mesmo nome de base são automaticamente combinados numa única amostra.

Ficheiros DRAGEN. Os ficheiros e formatos esperados para cada workset são os seguintes, quando os ficheiros VCF - Secondary Analysis data e os metadados Illumina são selecionados quando o projeto é criado:

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

Ficheiros Oxford Nanopore Technologies. Os ficheiros e formatos esperados são os seguintes ao selecionar VCF files - Secondary Analysis data, e metadados da Oxford Nanopore Technologies durante a criação do projeto. Os utilizadores podem importar um ficheiro de variante (SNP) por si só ou combiná-lo com ficheiros suplementares para análise CNV, STR ou SV:

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

Thermofisher/Affymetrix files. O formato de importação para ficheiros de análise secundária CGH é o formato .txt gerado pela aplicação Affymetrix da Thermofisher. Este formato é suportado apenas quando os metadados Thermofisher/Affymetrix e CGH MicroArrays são selecionados durante a criação do projeto. Os ficheiros e formatos esperados para cada workset são os seguintes:

GermVar Tertiary ARRAY_SampleID.txt

Ficheiros CGH MicroArrays. O formato de importação para ficheiros de análise secundária CGH é o formato .xls gerado pela aplicação Cytogenomics da Agilent. Este formato é suportado apenas quando os metadados Agilent e CGH MicroArrays são selecionados durante a criação do projeto. Os ficheiros e formatos esperados para cada workset são os seguintes:

GermVar Tertiary IntervalBasedReport_sample.xls

Ficheiros Alissa Reporter. O formato de importação dos ficheiros de análise secundária do Alissa Reporter é o formato .xls gerado pela aplicação Alissa Reporter da Agilent. Este formato é suportado apenas quando os metadados Agilent e CGH MicroArrays são selecionados durante a criação do projeto. Os ficheiros e formatos esperados para cada workset são os seguintes:

GermVar Tertiary IntervalBasedReport_sample.xls
ABERRATION_&_LOH_INTERVAL_sample.xls

PacBio-HiFi files. Ao criar um projeto, se forem escolhidos ficheiros VCF (Secondary Analysis data) e metadados PacBio-HiFi, os ficheiros e formatos esperados para cada workset são os seguintes:

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

Ficheiros VCF para painéis de genes somáticos. O formato de importação para ficheiros de análise secundária SNV, CNV e Fusions é o formato .vcf.gz gerado pelas aplicações Illumina, AmoyDx, Pillar e Thermofisher. Os nomes dos ficheiros originais têm de ser convertidos para corresponderem ao formato de nomenclatura necessário descrito abaixo.

Os ficheiros e formatos esperados são os seguintes quando se selecionam ficheiros VCF - dados da Secondary Analysis e metadados do painel de genes durante a criação do projeto somático. Os utilizadores podem importar um ficheiro de variantes (SNV) por si só ou combiná-lo com ficheiros suplementares para análise de CNV ou fusões:

SomaVar Tertiary

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

Os ficheiros e formatos esperados são os seguintes ao seleccionar «VCF files - Secondary Analysis data» e metadados «Comprehensive Genomic Profiling» durante a criação do projeto somático. Os utilizadores podem importar um ficheiro VCF multivariável, como o ficheiro VCF «All Variants» do Oncomine™ Comprehensive Assay V3/Plus.

SomaVar Tertiary

Oncomine_<LibPrepID>_<analysisID>.vcf

10.3. Metadados da amostra

Para cada amostra carregada, os utilizadores podem atribuir valores aos metadados que seleccionaram ao criar o projeto ou definir novos valores clicando no ícone no canto superior direito da lista de amostras . .

Alguns metadados de amostra podem ser carregados diretamente como um ficheiro de valores separados por separadores, seleccionando o ícone  .

Estes critérios desempenham um papel importante na garantia da rastreabilidade dos dados e na referência cruzada de diferentes análises efectuadas na plataforma. Alguns metadados, como o género, a doença e as identificações HPO, estão envolvidos no processo de identificação de variantes.

Quando o campo tipo de ensaio («Assay Type») é preenchido com «Germline», os metadados associados à amostra por predefinição são «HPO», «Sex» e grupo «GnomAD». Quando o campo tipo de ensaio («Assay Type») é preenchido com «Somatic», os metadados associados à amostra por predefinição são «Disease» e «Sex».

Para adicionar ou modificar HPOs numa amostra, o utilizador pode clicar em "+ Add" na coluna HPO depois de seleccionar uma amostra. Isto abre uma janela pop-up que permite adicionar HPOs um a um, procurando um fenotipo ou código HPO na barra de pesquisa. É igualmente

possível importar um ficheiro de texto que contenha códigos HPO, clicando no ícone  . Está então disponível um resumo dos termos introduzidos, com a opção de eliminar termos, se necessário.

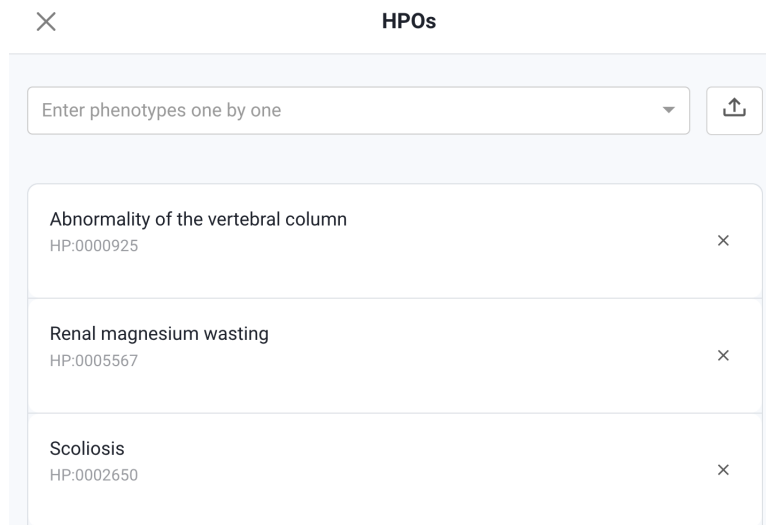


Figura 6: Janela pop-up de importação de HPO

10.4. Editar uma amostra

O nome, bem como os metadados associados a uma amostra, podem ser editados pelo utilizador em qualquer altura, seleccionando a amostra na lista "Samples" e clicando no ícone



. As modificações devem então ser validadas para serem efectivas.

10.5. Eliminar uma amostra

Os utilizadores podem eliminar os ficheiros FastQ de uma amostra da plataforma clicando no ícone  depois de seleccionar a amostra pretendida. Estes ficheiros libertarão espaço de armazenamento e os resultados continuarão a estar disponíveis no SeqOne.

Para apagar uma amostra, clicar no ícone  no separador "Samples" depois de seleccionar a amostra pretendida. A amostra só pode ser removida permanentemente após a eliminação de todas as análises a ela associadas.

11. Análise

11.1. Lançamento de uma nova análise

Os utilizadores podem lançar uma nova análise clicando em "New analysis" no ecrã *Overview* de cada projeto.

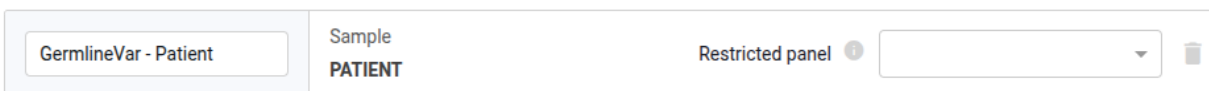
O lançamento da análise é efectuado em três etapas distintas, descritas a seguir.

O passo "1. Workset" permite-lhe escolher o workset de análise apropriado (ver secção **Escolher o workset de análise**).

A etapa "2. Select Samples" está dividida em duas partes:

- 2.1. *Set options*:
 - descreve o workset selecionado e o genoma de referência
 - permite adicionar *preenchimento* através da funcionalidade "Extra calling padding", se aplicável (ver secção **Escolher o workset de análise**)
 - permite escolher o modo de processamento UMI (ver secção **Escolher o modo de processamento UMI**), se aplicável.
 - permite escolher a linha de base CNV (baseline CNV), a linha de base MSI (baseline MSI) ou ajustar o limite mínimo de VAF, se aplicável.
 - Permite escolher o protocolo.
- 2.2. *Select sample(s)*: permite selecionar amostras para análise. Os utilizadores podem selecionar especificamente novas amostras, que estão marcadas como "New", utilizando a função "select all new".

O passo "3. Confirm Analysis" resume as amostras selecionadas. Para cada análise a lançar, é possível aplicar um painel restrito selecionando um ou mais painéis de genes através da opção "Restricted panel" (ver também a secção **Vista modal Panel**).



A opção "Restricted panel" aplicará um filtro à análise sob a forma de um ou mais painéis de genes. Esta opção não afecta a chamada de variantes, que é determinada pelo manifesto do projeto e pelo *preenchimento adicional*. A escolha de "Restricted panel" é reversível; o utilizador pode modificar estes painéis a partir do separador *Panel (Painel)* do painel lateral de gestão de análises após a análise (ver secção **Vista modal Panel**).

11.1.1. Escolher o workset de análise

Estão disponíveis vários "worksets" de análise para responder a várias questões biológicas. Os worksets propostos aquando da criação de uma nova análise dependem dos metadados definidos aquando da criação do projeto e podem, portanto, variar em função dos dados.

Worksets de análise disponíveis a partir de ficheiros FASTQ

Os seguintes worksets estão disponíveis quando os metadados *Fastq files - Primary Analysis data* são selecionados durante a criação do projeto.

Excluindo os projetos em que o método de amplificação é o "Amplification", alguns worksets" (ou seja, GermlineVar, GermlineFamily, SomaVar, SomaVar LF (versões ≥ 2.0) e SomaCGP (versão ≥ 0.6), GermVar, GermVar Family, SomaLBx, SomaMethyl) permitem alargar a área de pesquisa para a identificação de variantes e adicionar entre 0 e 100 pb de preenchimento ao manifesto 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.

A opção está disponível na etapa 2 (Select Samples) do lançamento da análise através da função "Extra calling padding". Por defeito, o valor é definido como 0 e tem de ser um número inteiro natural.

Alguns worksets (ou seja, SomaVar (versões ≥ 2.6) e SomaHemato (versões ≥ 0.2)) oferecem uma opção para ajustar o limite mínimo de VAF para deteção de variantes. Esse limite pode ser definido entre 0,01% e 1%, sendo 1% o valor padrão.

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 obter mais informações sobre essa opção, entre em contacto com o suporte ao utilizador em support@seqone.com.

GermlineVar



GermlineVar é utilizado para detetar mutações germinativas com base na sequenciação de um painel de genes ou do exoma do paciente.

Ao executar uma análise GermlineVar, os utilizadores são convidados a:

- Selecionar a versão do workset a executar e clicar em "Select",
- Selecionar as amostras a analisar e, em seguida, clicar em "Next step",
- Iniciar a análise clicando em "Start analysis".

GermVar RUO



GermVar permite a deteção de mutações germinativas, com base na sequenciação de um painel de genes, do exoma ou do genoma de um paciente

Ao executar uma análise GermVar, os utilizadores são convidados a:

- Selecionar a versão do workset que pretende executar e clicar em "Select",
- Selecione o modelo a ser utilizado,
- Selecione a(s) amostra(s) a analisar e clique em "Next step",
- Iniciar a análise clicando em "Start analysis".

GermlineVarLR



GermlineVarLR é utilizado para detetar mutações da linha germinativa em dados de sequenciação de leitura longa de sequenciação de painéis de genes.

Ao executar uma análise GermlineVarLR, os utilizadores são convidados a:

- Selecionar a versão do workset a executar e clicar em "Select",
- Selecionar o modelo a utilizar e a(s) amostra(s) a analisar e, em seguida, clicar em "Next step",
- Iniciar a análise clicando em "Start analysis".

GermlineFamily



O GermlineFamily é utilizado para detetar mutações hereditárias com base na sequenciação de um painel de genes, do exoma ou do genoma de um paciente e dos membros da sua família. Este workset também pode ser utilizado para "Carrier screening".

Ao executar uma análise GermlineFamily, os utilizadores devem:

- Selecionar a versão do workset a executar e, em seguida, clicar em "Select",
- Selecionar pelo menos 2 amostras,
- Na secção "Set conditions", indicar se a amostra é do paciente ou de um membro da família, definindo o vínculo familiar em "Family link". O paciente é automaticamente marcado como "affected" (afetado). No caso de um "Carrier screening", indicar para cada amostra a relação familiar por "Mother" ou "Father" sem marcar a caixa "Is affected?",
- Verificar se os parâmetros coincidem e, em seguida, clicar em "Next step",
- Iniciar a análise clicando em "Start analysis".

GermVar Family



GermVar Family permite a deteção de mutações germinativas, com base na sequenciação de um painel de genes, do exoma ou do genoma de um paciente

Ao executar uma análise GermVar Family, os utilizadores são convidados a:

- Selecionar a versão do workset a executar e, em seguida, clicar em "Select",
- Selecione o modelo a ser utilizado,
- Selecionar pelo menos 2 amostras. Selecionar pelo menos 2 amostras,
- Na secção "Set conditions", indicar se a amostra é do paciente ou de um membro da família, definindo o vínculo familiar em "Family link". O paciente é automaticamente marcado como "affected" (afetado). No caso de um "Carrier screening", indicar para cada amostra a relação familiar por "Mother" ou "Father" sem marcar a caixa "Is affected?",
- Verificar se os parâmetros coincidem e, em seguida, clicar em "Next step",
- Iniciar a análise clicando em "Start analysis".

CNVCapture



O CNVCapture é utilizado para detetar variações do número de cópias da linha germinal (CNVs) a partir de dados de captura de painéis de genes.

Ao executar uma análise CNVCapture, os utilizadores são convidados a:

- Selecionar a versão do workset a executar e, em seguida, clicar em "Select",
- Selecionar um número mínimo de 8 amostras a analisar e clicar em "Next step",
- Executar a análise clicando em "Start analysis".

CNVExome



O CNVExome é utilizado para detetar variações do número de cópias (CNVs) a partir de dados de captura do exoma.

Ao executar uma análise do CNVExome, os utilizadores são convidados a:

- Selecionar a versão do workset a executar e clicar em "Select",
- Selecionar o modelo a utilizar e a(s) amostra(s) a analisar e, em seguida, clicar em "Next step",
- Iniciar a análise clicando em "Start analysis".

SomaVar



SomaVar detecta mutações somáticas em amostras de tumores. Este workset pode ser utilizado com dados de captura e de amplicon.

Ao executar uma análise SomaVar, os utilizadores são convidados a:

- Selecionar a versão do workset a executar e clicar em "Select",
- Nas opções, selecionar o modo de processamento UMI aplicável a partir dos modos elegíveis para a análise (ver secção **Escolher o modo de processamento UMI**),
- Selecionar a(s) amostra(s) a analisar e clicar em "Next step",
- Iniciar a análise clicando em "Start analysis".

SomaCNVCapture



O SomaCNVCapture é utilizado para detetar variações somáticas do número de cópias (CNVs) a partir de dados de captura de painéis de genes.

Ao executar uma análise SomaCNVCapture, os utilizadores são convidados a:

- Selecionar a versão do workset a executar e clicar em "Select",
- Nas opções, seleccionar o modo de processamento UMI aplicável a partir dos modos elegíveis para a análise (ver secção **Escolher o modo de processamento UMI**),
- Selecionar um número mínimo de oito amostras a analisar e, em seguida, clicar em "Next step",
- Executar a análise clicando em "Start analysis".

SomaRNA



O SomaRNA pode ser utilizado para detetar fusões de genes a partir de dados de captura de RNA-seq.

Ao executar uma análise SomaRNA, os utilizadores são convidados a:

- Selecionar a versão do workset a executar e clicar em "Select",
- Nas opções, seleccionar o modo de processamento UMI aplicável a partir dos modos elegíveis para a análise (ver secção **Escolher o modo de processamento UMI**),
- seleccionar a(s) amostra(s) a analisar e, em seguida, clicar em "Next step",
- Executar a análise clicando em "Start analysis".

SomaMSI



O SomaMSI detecta o estado das regiões de microssatélites (MSS ou MSI).

Ao executar uma análise SomaMSI, os utilizadores são convidados a:

- Selecionar um número mínimo de oito amostras a analisar e, em seguida, clicar em "Next step",
- Nas opções, seleccionar o modo de processamento UMI aplicável a partir dos modos elegíveis para a análise (ver secção **Escolher o modo de processamento UMI**),

- selecionar um número mínimo de 8 amostras a analisar e clicar em "Next step",
- Executar a análise clicando em "*Start analysis*".

SomaHRD

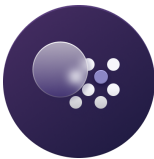


O SomaHRD permite a deteção do status HRD a partir de dados WGS superficiais isolados ou acompanhados por um painel de genes incluindo BRCA1 e BRCA2.

Ao executar uma análise SomaHRD, os utilizadores são convidados a:

- Selecionar a versão do workset a executar e clicar em "*Select*",
- Introduzir os metadados de cellularity (% de células tumorais) para cada amostra (ver secção **Sample metadata**),
- Selecionar a(s) amostra(s) a analisar e clicar em "*Next step*",
- Selecionar a análise SomaVar ou SomaCGP (GRCh37 ou GRCh38) adequada para cada amostra utilizando a opção "*Select panel analysis (Optional)*",
- Executar a análise clicando em "*Start analysis*".

SomaVar LF

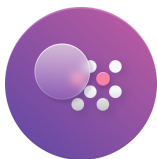


Deteção de variantes curtas de baixa frequência e de variantes estruturais, bem como de genes de fusão, a partir de amostras de tumores (biópsias sólidas ou líquidas, ADN tumoral circulante) enriquecidas por captura num painel de genes.

Ao executar uma análise SomaVar LF, os utilizadores são convidados a:

- Selecionar a versão do workset a executar e clicar em "*Select*",
- Nas opções, selecionar o modo de processamento UMI aplicável a partir dos modos elegíveis para a análise (ver secção **Escolher o modo de processamento UMI**),
- Selecionar a(s) amostra(s) a analisar e clicar em "*Next step*",
- Executar a análise clicando em "*Start analysis*".

SomaCGP



O SomaCGP permite a deteção de mutações somáticas, variantes estruturais, fusões de genes, bem como o estado MSI e TMB a partir de amostras de tumores. Este workset é dedicado à análise de ADN FFPE preparado com o painel SureSelect Cancer CGP Assay da Agilent e só está disponível quando os metadados Comprehensive Genomic Profiling são selecionados na criação do projeto. Este workset pode ser utilizado com dados de captura.

Ao executar uma análise SomaCGP, os utilizadores são convidados a:

- Selecionar a versão do workset a executar e clicar em "Select",
- Se solicitado, selecione uma «CNV baseline» e/ou «MSI baseline» (se nenhuma linha de base for selecionada, a análise será executada sem deteção CNV/MSI),
- Nas opções, selecionar o modo de processamento UMI aplicável a partir dos modos elegíveis para a análise (ver secção **Escolher o modo de processamento UMI**),
- Selecionar a(s) amostra(s) a analisar e clicar em "Next step",
- Iniciar a análise clicando em "Start analysis".

SomaHemato



O SomaHemato permite a deteção de mutações somáticas a partir de amostras com especificidades hematológicas. Este workset pode ser utilizado com dados de captura.

Ao executar uma análise SomaHemato, os utilizadores são convidados a:

- Selecionar a versão do workset a executar e clicar em "Select",
- Nas opções, selecionar o modo de processamento UMI aplicável a partir dos modos elegíveis para a análise (ver secção **Escolher o modo de processamento UMI**),
- Selecionar a(s) amostra(s) a analisar e clicar em "Next step",
- Iniciar a análise clicando em "Start analysis".

SomaLBx



O SomaLBx permite a deteção de mutações somáticas de baixa frequência, variantes estruturais, fusões de genes, a partir de amostras de tumores. Este workset é dedicado à análise de ADNct e pode ser utilizado com dados de captura.

Ao executar uma análise SomaLBx, os utilizadores são convidados a:

- Selecionar a versão do workset a executar e clicar em "Select",
- Selecione uma «CNV baseline» (se nenhuma linha de base for selecionada, a análise será executada sem deteção de CNV),
- Nas opções, seleccionar o modo de processamento UMI aplicável a partir dos modos elegíveis para a análise (ver secção **Escolher o modo de processamento UMI**),
- Selecionar a(s) amostra(s) a analisar e clicar em "Next step",
- Iniciar a análise clicando em "Start analysis".

SomaMethyl



O SomaMethyl permite a deteção de ilhas CpG metiladas a partir de amostras de ADN tumoral circulante. Este conjunto de trabalho é dedicado à análise da metilação no ctDNA e pode ser utilizado com dados de captura de bisulfato.

Ao executar uma análise SomaMethyl, os utilizadores são convidados a:

- Selecionar a versão do conjunto de trabalho a ser executada e clicar em «Select».
- Selecionar uma «Methyl baseline» correspondente à coorte de controlo (se nenhuma coorte for selecionada, a análise será executada sem deteção de limiar e metilação diferencial),
- Selecionar a(s) amostra(s) a ser(em) analisada(s) e, em seguida, clicar em «Next step»,
- Iniciar a análise clicando em «Start analysis».

Worksets de análise disponíveis a partir de ficheiros de análises secundárias

Os seguintes worksets estão disponíveis quando os *dados VCF files - Secondary Analysis* são selecionados durante a criação do projeto ou quando os metadados Agilent e CGH MicroArrays são selecionados (ver secção **Ficheiros de análise secundária**). Neste caso, a plataforma recebe dados de saída da aplicação do fornecedor selecionado, procede à anotação e injecta variantes nas respectivas interfaces de visualização.

GermVar Tertiary



Pipeline de análise de amostra única de variantes detetadas pela aplicação Illumina Germline ou pela aplicação EPI2ME da Oxford Nanopore Technologies, a partir de dados WGS e WES ou variantes detetadas pela aplicação Cytogenomics da Agilent a partir de dados CGH MicroArrays.

Ao executar uma análise GermVar Tertiary, os utilizadores são convidados a:

- Selecionar a versão do workset a executar e, em seguida, clicar em "Select",
- Selecionar a(s) amostra(s) a analisar e clicar em "Next Step".
- Iniciar a análise clicando em "Start analysis".

GermVar Tertiary Family

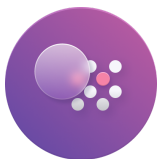


Pipeline de análise multi-sample de variantes detetadas pela aplicação Illumina Germline ou pela aplicação EPI2ME da Oxford Nanopore Technologies, a partir de dados WGS e WES.

Ao executar uma análise GermVar Tertiary Family, os utilizadores são convidados a:

- Selecionar a versão do workset a executar e, em seguida, clicar em "Select",
- Selecionar pelo menos 2 amostras,
- Na secção "Set conditions", indicar se a amostra é do paciente ou de um membro da família, definindo o vínculo familiar em "Family link". O paciente é automaticamente marcado como "affected" (afetado).
- Verificar se as definições coincidem.
- Executar a análise clicando em "Start analysis".

SomaVar Tertiary



O SomaVar Tertiary é um pipeline de análise terciária baseado em VCF que processa ficheiros VCF Illumina WES e ficheiros VCF de painéis de genes de vários fornecedores de kits (ver secção **Formato do ficheiro de entrada**).

Ao executar uma análise SomaVar Tertiary, os utilizadores são convidados a:

- Selecionar a versão do workset a executar e, em seguida, clicar em "Select",
- Selecionar a(s) amostra(s) a analisar e clicar em "Next Step".
- Iniciar a análise clicando em "Start analysis".

11.1.2. Escolher o modo de processamento de UMI

Apresentação dos modos de processamento de UMI

Na prática, a plataforma propõe um ou vários métodos de tratamento das UMI aquando da execução de uma análise:

1. **Standard mode** (recomendado): os consensos são gerados a partir dos duplicados de PCR que apresentam a mesma UMI quando o seu número é maior ou igual a 2. As UMIs representadas por uma única sequência (singletons) são também retidas. A frequência alélica mínima das variantes comunicadas e o número mínimo de reads que suportam a mutação (AO, Alternate Observation) diferem consoante o workset utilizado:
 - Em SomaVar, SomaHemato, SomaRNA, SomaCNVcapture, SomaCGP e SomaMSI, a frequência alélica mínima é de 1% e o AO é um mínimo de 3 reads.
 - Em SomaVar LF e SomaLBx, a frequência alélica mínima é de 0,1% e o AO mínimo é de 2 reads. Este workset permite a deteção de variantes com uma frequência alélica inferior a 1%. É recomendado quando a profundidade de sequenciação é superior a 5000x e para aplicações como a sequenciação de ADNct.
2. **UMI disabled** (disponível apenas em SomaVar, SomaRNA, SomaHemato, SomaLBx, SomaCNVcapture, SomaCGP e SomaMSI): Os UMIs não são utilizados para deduplicação e são cortados do fim das sequências antes da análise. Por conseguinte, os dados gerados por este pipeline de análise são brutos (não deduplicados). A frequência alélica mínima das variantes registadas é de 1% e o número mínimo de reads que suportam a mutação (AO) é de 3.

As diferenças nas definições entre estes modos estão resumidas na tabela seguinte:

	UMI disabled	UMI standard
Qualidade mínima mediana/média da base mutada (Phred score)*	30	30
Número mínimo de reads por consenso	N/A	2
Qualidade mínima de cada base no consenso (pontuação Phred)	N/A	30
Read sem consensos filtrados	N/A	não

Quadro 1: Settings das diferentes configurações de processamento UMI

* Estes filtros aplicam-se apenas a variantes que não são selecionadas pelo modelo SVM após a chamada de variantes (pipelines SomaVar e SomaRNA).

Processo de tratamento de UMI

A estrutura dos UMIs depende de cada kit do fornecedor e os seguintes kits estão atualmente disponíveis na 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: A configuração UMI "QIAseq Targeted DNA Pro" deve ser utilizada apenas para SomaVar, com um comprimento de leitura de pelo menos 150 bp. Para mais informações, contacte o suporte.

O processamento de UMIs baseia-se na utilização do conjunto de ferramentas fgbio, que pode tratar UMIs e reads marcadas com UMIs.

O nosso pipeline de análise UMI consiste nas seguintes etapas:

1. Se necessário, cortar os UMIs do final das sequências e gerar um BAM não alinhado a partir do par de ficheiros FASTQ,
2. Alinhar sequências sem UMI com BWA-MEM2 (ADN/ctADN) ou STAR (ARN)
3. Fundir os ficheiros BAM que contêm os UMIs com as sequências alinhadas utilizando o ZipperBam,
4. Agrupar reads se tiverem a mesma posição e partilharem o mesmo UMI,
5. Gerar a sequência de consenso para cada grupo de UMIs,
6. Filtrar cada base de acordo com a sua qualidade: a confiança na sequência de consenso aumenta com o número de reads por grupo,
7. Filtrar sequências de consenso: se mais de 20% das bases da sequência de consenso forem filtradas, a sequência de consenso não é mantida,
8. Conversão destas sequências de consenso filtradas em FASTQ e realinhamento final no genoma de referência com BWA-MEM2 (ADN/ctADN) ou STAR (ARN).

Compatibilidade dos worksets de análise

Os "worksets" SomaVar, SomaHemato, SomaCGP, SomaLBx, SomaMethyl, SomaVar LF, SomaCNVCapture", SomaMSI e SomaRNA são compatíveis com os dados UMI. As opções de processamento disponíveis aquando do lançamento dos diferentes worksets são descritas na tabela abaixo:

	UMI disabled	UMI standard
SomaVar	Sim	Sim

SomaCNVCapture".	Sim	Sim
SomaMSI	Sim	Sim
SomaRNA	Sim	Sim
SomaVar LF	Não	Sim
SomaLBx	Sim	Sim
SomaHemato	Sim	Sim
SomaMethyl	Não	Sim

Tabela 2: Opções de processamento UMI disponíveis aquando do lançamento dos worksets compatíveis.

11.2. Estado da análise

Depois de lançar uma análise, a plataforma SeqOne utiliza diferentes status para permitir aos utilizadores acompanhar o progresso da sua análise:

- **Bioinformatics Pending:** a plataforma está à espera que outra análise termine
- **Bioinformatics Running:** a análise está em curso; este status mostra também qual a fase da análise que está a ser realizada
- **Bioinformatics Finished:** a análise foi concluída
- **Analysis failed** - error: ocorreu um erro durante a análise; neste caso, os utilizadores são convidados a contactar a equipa de apoio ao utilizador
- **Interpretation - new:** a análise está terminada e pronta para interpretação, ainda não foi aberta por nenhum utilizador
- **Analysis failed - error:** A análise deparou-se com um erro. Neste caso, o utilizador é convidado a contactar o apoio ao utilizador.
- **Analysis failed - canceled":** a análise não foi iniciada porque a amostra a ela ligada encontrou um erro durante o carregamento ou a verificação.

11.3. Qualidade da amostra

O separador QC do projeto está dividido em 2 sub-fichas: 'General statistics' e 'ID check summary'.

[General statistics'](#). Este separador apresenta uma tabela que representa a qualidade geral do alinhamento e da cobertura para cada amostra do projeto. O utilizador pode exportar a tabela como um ficheiro TSV, clicando em .

O cálculo baseia-se na ferramenta SeQC, um programa desenvolvido pela SeqOne para calcular as estatísticas de alinhamento e cobertura de todas as reads alinhadas com o genoma.

O utilizador pode optar por utilizar as métricas QC predefinidas ou definir limiares personalizados a partir do separador "Settings" do projeto (ver secção **Configurar métricas QC predefinidas do projeto**).

As métricas propostas são as seguintes:

- número total de reads
- comprimento médio dos reads
- tamanho médio dos fragmentos de ADN na biblioteca (inserções)
- percentagem de reads de qualidade >30
- percentagem de reads alinhadas com o genoma de referência
- percentagem de reads no objetivo
- percentagem de reads duplicados
- taxa de transição/transversão
- Cobertura média
- Fração de alvos abrangidos em diferentes limiares, dependendo do contexto (amplicon, Capture).

Os alvos do painel ou do exoma são determinados a partir do ficheiro BED selecionado quando o projeto é criado.

ID check summary. Este separador detalha os genótipos de todos os marcadores incluídos nas suas análises para cada amostra (ver parágrafo **Marcador** na secção **Managing user parameters**). Uma linha corresponde a uma amostra e cada coluna representa um marcador com o respetivo ID associado. O utilizador pode exportar esta tabela como um ficheiro TSV clicando em .

12. Controlo de qualidade dos dados

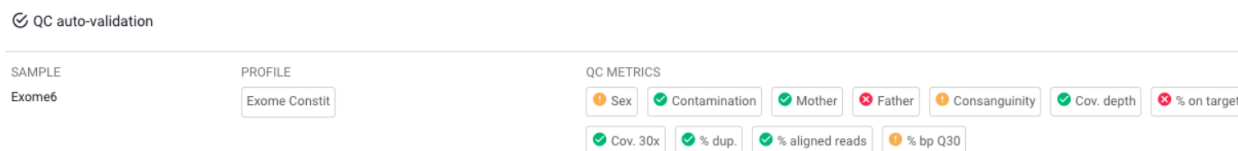
12.1. Métricas e relatórios QC

12.1.1. Métricas de QC e controlos de QC

A partir do separador *QC*", o utilizador acede aos controlos de qualidade aplicados à análise, sob a forma de sub-tabs descritas nas secções abaixo: *QC report*", *ID check*" (se aplicável) e *Technical report*".

No caso de uma análise CGH MicroArrays, o separador *QC*" não está disponível. No entanto, o relatório técnico está disponível no separador *Files (Ficheiros)* da análise em formato *report.html*".

Na parte superior do separador *QC*, uma secção "*QC auto-validation*" permite ao utilizador codificar por cores as métricas de QC como *Passed*, *Warning* ou *Failed*. Ao clicar no ícone associado, o utilizador acede aos limiares aplicados e aos detalhes do QC: sexo, contaminação, parentesco e consanguinidade, profundidade de cobertura, % no alvo, cobertura 30X ou 100X, % duplicados, % reads alinhadas com o genoma de referência ou % bp Q30.



12.1.2. Relatório QC

O *relatório global QC* do processo de sequenciação contém várias métricas que descrevem a qualidade e o alinhamento das sequências, a cobertura das regiões no manifesto e algumas estatísticas relacionadas com as variantes. Este relatório é composto pelas seguintes secções:

[Alignment details \(Detalhes do alinhamento\)](#) (SomaMethyl). Total de leituras processadas, leituras alinhadas de forma única e não única, e taxas de incompatibilidade.

[Deduplication \(Deduplicação\)](#) (SomaMethyl). Métricas de duplicação para leituras alinhadas, destacando problemas potenciais com a preparação da biblioteca.

[Cytosine coverage & methylation \(Cobertura de citosina e metilação\)](#) (SomaMethyl). Taxas de conversão de citosina para uracilo para cada contexto (CpG, CHG, CHH).

[Estatísticas de grupos UMI](#) (apenas SomaVar, SomaVar LF, SomaMethyl e SomaRNA UMI). Consulte o parágrafo **Escolher o modo de processamento UMI**.

O relatório QC das análises SomaVar, SomaVar LF, SomaMethyl e SomaRNA oferece uma visão mais pormenorizada da composição da amostra, em particular a distribuição das UMIs de acordo com o número de sequências que as albergam.

As estatísticas descrevem em particular:

- A fração de singletons (UMI transportados por uma sequência única)
- A dimensão média e mediana dos grupos UMI, bem como o 'standard deviation'
- O tamanho do maior grupo e a fração do total de sequências que inclui.

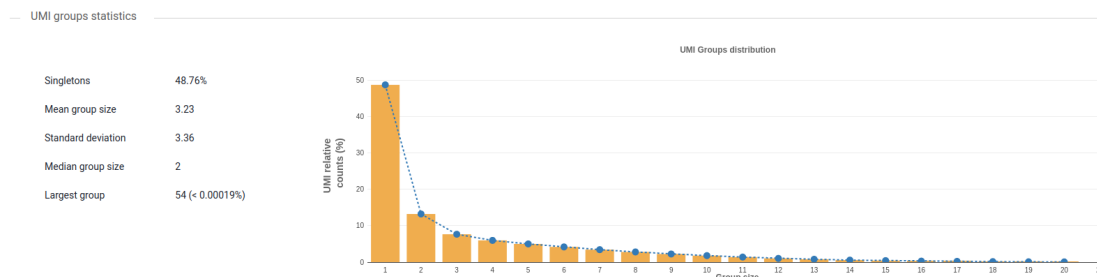


Figura 7: Estatísticas relativas à distribuição UMI em termos do número de sequências numa amostra.

As métricas globais de "QC" para o projeto são também mais detalhadas e divididas em três secções:

- Métricas QC após o processamento UMI, que indicam a tendência em cada métrica (em percentagem) com base nos dados em bruto.
- 'Raw QC metrics' calculadas a partir de todas as sequências (antes do processamento UMI).
- Distribuição da UMI em termos do número de sequências em cada amostra.

Cross-contamination estimate' (apenas SomaVar LF). A estimativa da contaminação cruzada é efectuada individualmente com a ferramenta 'CalculateContamination' do GATK'. Esta última utiliza o sinal dos alelos de referência detectados em loci homozigóticos para calcular a fração de sequências no indivíduo provenientes de contaminação cruzada.

Sample integrity and identity (apenas análises da linha germinal). Esta secção enumera métricas relacionadas com a integridade e identidade da amostra:

- **Family link**": metadados definidos pelo utilizador
- **Original sex**": metadados relativos ao sexo introduzidos pelo utilizador (Male/Female/Unknown)
- **Detected sex**: sexo detectado (Male/Female/Unknown)

- **Relationship coefficient** : coeficiente de relação calculado utilizando a ferramenta *Somalier* para cada amostra na análise da família, esta métrica apresenta 'N/A' para a própria amostra.
- **Contamination**: percentagem estimada de contaminação da amostra por ADN estranho, utilizando a ferramenta 'verifyBamID2'.

São também listados o número de variantes homozigóticas e heterozigóticas, o rácio destes dois valores (rácio Hom/Het), o número de variantes homozigóticas e heterozigóticas no cromossoma X, o rácio equivalente no cromossoma X (rácio Het/Hom no chr X') e o número de variantes no cromossoma Y.

As métricas *Family link* e *Relationship coefficient*, utilizadas para verificar a relação entre os indivíduos de uma família, só são apresentadas em análises de famílias trio (GermlineVar Family ou GermVar Tertiary Family).

As métricas de género são utilizadas para verificar a correspondência entre o género introduzido e o género detectado nos dados.

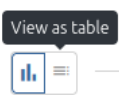
A métrica de contaminação não é calculada para os worksets GermVar Tertiary e GermVar Tertiary Family.

As verificações QC sobre correspondência de género, contaminação e relação estão disponíveis ao nível do projeto ou da análise para identificar facilmente quaisquer correspondências ou incompatibilidades.

Alignment metrics'. Esta secção lista as principais métricas calculadas no final do alinhamento, tais como:

- O número total de sequências na amostra,
- O tamanho mediano das inserções, calculado a partir da distância entre os dois reads do mesmo par, cujo respetivo comprimento está incluído,
- O comprimento médio da sequência,
- A fração de sequências alinhadas com sucesso com o genoma de referência,
- A taxa de duplicação (apenas Capture),
- A fração de bases cuja qualidade é superior a 30,
- A cobertura média das regiões-alvo, determinada a partir do manifesto de análise.

Coverage. Gráfico que mostra a fração de regiões-alvo abrangidas em diferentes limiares, indicados no eixo x. Os valores de cobertura podem ser apresentados sob a forma de tabela,

clicando no botão  que permite ao utilizador copiar/colar esses valores.

QC Graphs. Gráficos da análise de qualidade FastQC, ilustrando em particular o conteúdo GC das sequências, bem como a qualidade das bases de acordo com a sua posição nas sequências.

Variant statistics. Uma lista de estatísticas sobre as variantes detetadas na análise, que incluem, nomeadamente:

- O número de transições (Ts) e transversões (Tv) entre as substituições detetadas,
- O rácio Ts / Tv,
- O número total de mutações detetadas,
- A fração de SNVs, MNVs e indels neste total.

12.1.3. Relatório de monitorização da identidade 'ID-Check'.

O separador 'ID check' é dedicado aos resultados da monitorização da identidade no separador QC dos worksets que chamam variantes.

Os resultados são apresentados em forma de tabela com as seguintes informações:

- Uma descrição de cada marcador com base em coordenadas genómicas, por exemplo: chr-start-ref-alt,
- uma etiqueta (rsID, etiqueta personalizada) quando dada pelo utilizador,
- genótipo observado (GT) ou frequência alélica (VAF),
- a profundidade de cobertura na localização do marcador (COV),
- a qualidade da base de dados (QUAL).

Quando são incluídas várias amostras na análise, as colunas específicas das amostras (GT/VAF, QUAL, COV) são repetidas para cada amostra individualmente.

A seleção de uma variante na tabela abre um navegador IGV num novo separador, centrado nas coordenadas genómicas da variante.

12.1.4. Relatório técnico

O *relatório técnico* enumera as ferramentas utilizadas durante a análise e a sua versão. Pode ser acedido no separador QC de uma análise e depois em "Technical report".

Em particular, descreve :

- informações gerais sobre a configuração e os metadados do projeto e da análise, tais como o tipo de sequenciação, a versão do genoma e a versão do workset utilizado;

[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

- pormenores do manifesto utilizado, como o nome, o Size, o comprimento (bp) e o padding adicionado à análise;
- pormenores sobre as bases de dados e as ferramentas bioinformáticas utilizadas.

12.1.5. Relatório UPD

O Relatório UPD está disponível na seção QC de análises germinativas compatíveis (por ex., GermVar e fluxos de trabalho relacionados). Ele resume os resultados da detecção de UPD e fornece elementos visuais para apoiar a interpretação.

O Relatório UPD normalmente inclui:

- Visão geral dos resultados: indicação de se a análise detectou um evento do tipo UPD e o tipo de evento (por ex., completo ou segmentar; em análises de trio, o relatório também pode apoiar a origem parental e os subtipos de eventos, como UPD | Iso / UPD | Het).
- Lista de eventos: uma lista de eventos detectados com sua localização cromossômica e tamanho, e a(s) pontuação(ões) associada(s) usada(s) pelo algoritmo.
- Gráficos de visualização: gráficos baseados em genótipos SNP (por ex., frequência do alelo B (BAF) e sinal de homozigidade/ROH) que ajudam a confirmar a presença de longas regiões homozigóticas e seus limites.

Informações de controle de qualidade: métricas relacionadas à confiabilidade da detecção (por ex., densidade de SNP / número de SNPs) e avisos quando os dados não são suficientes para uma chamada confiável.

Os eventos UPD detectados também são exibidos no Large Variant Viewer (LVV) com rótulos específicos (por ex., UPD | Iso e UPD | Het).

Pontuação UPD

A pontuação UPD é um indicador numérico usado para quantificar com que força a amostra mostra um sinal do tipo UPD ao longo do genoma. A detecção de UPD no SeqOne depende dos padrões de genótipo SNP:

- Em uma análise singleton (solo), o algoritmo procura grandes regiões de homozigidade (ROH) e por mudanças abruptas na distribuição de SNPs ao longo dos cromossomos.
- Em uma análise de trio, o padrão SNP é usado adicionalmente para determinar a origem parental do evento e para distinguir isodissomia (UPD | Iso) de heterodissomia (UPD | Het).

Na prática, a pontuação UPD reflete a força e a consistência do sinal que suporta um evento UPD (por ex., padrão claro de ROH/BAF e limites bem delimitados). Uma pontuação mais alta indica que é mais provável que o evento seja um verdadeiro evento UPD, enquanto valores mais baixos indicam sinais mais fracos ou ruidosos.

A pontuação UPD deve sempre ser interpretada juntamente com os elementos fornecidos na aba Relatório UPD (seção QC), incluindo:

- os gráficos de visualização UPD (por ex., frequência do alelo B / sinal de homoziguidade),
- o tamanho e a localização cromossômica do evento,
- as métricas de qualidade da análise (por ex., densidade de SNP / número de SNPs).

Gráfico de frequência do alelo B (BAF) do relatório UPD

O gráfico de frequência do alelo B (BAF) é uma visualização dos genótipos SNP ao longo do genoma e é usado para avaliar o desequilíbrio alélico e a perda de homoziguidade (LOH).

Para cada posição SNP, a BAF corresponde à fração de leituras de sequenciamento que suportam o alelo "B" e varia de 0 a 1:

- Valores próximos a 0 correspondem a um genótipo homozigoto (AA).
- Valores próximos a 0,5 correspondem a um genótipo heterozigoto (AB).
- Valores próximos a 1 correspondem a um genótipo homozigoto (BB).

Em uma região diploide normal, o gráfico BAF normalmente mostra três bandas em torno de 0, 0,5 e 1. Em longas regiões de homoziguidade (ROH) ou isodissomia UPD, a banda heterozigota (em torno de 0,5) é reduzida ou ausente, o que suporta a presença de um sinal do tipo UPD.

Limitações importantes:

- Um sinal do tipo UPD também pode ser observado no contexto de consanguinidade parental (várias regiões ROH ao longo do genoma).
- A chamada de UPD é baseada principalmente na distribuição de SNPs e pode não ser suficiente sozinha para distinguir alguns padrões de UPD de eventos relacionados ao número de cópias; quando necessário, cruze os dados com outras evidências disponíveis (por ex., cobertura / resultados CNV).

13. Interpretação dos resultados da análise

13.1. Resultados da análise

Os utilizadores podem aceder às análises de um projeto clicando no ícone "Analyses" na caixa associada ou clicando no ícone com o mesmo nome no canto superior esquerdo do painel de controlo do utilizador.

Quando uma análise tiver sido concluída e estiver disponível para visualização, é apresentada com o estado "*Interpretation - New*".

Os resultados da análise estão então disponíveis sob a forma de vários relatórios, que podem ser abertos diretamente clicando neles.

Existem várias variações do estado "*Interpretation*", consoante os progressos alcançados na interpretação dos resultados:

- **New** - o quadro de variantes ainda não foi visualizado
- **In Progress** - a tabela de variantes foi visualizada, mas ainda não foram validadas quaisquer variantes
- **Evaluating** - pelo menos uma variante foi validada
- **Reporting validated** - foi gerado um relatório para a análise

13.1.1. Gestão do fluxo de trabalho de interpretação das análises

13.1.1.1. Atribuição de análises para interpretação

Os administradores da entidade podem optar por atribuir uma ou mais análises para serem interpretadas por diferentes utilizadores da entidade a partir do separador "Analyses" do projeto. Para tal, o administrador pode selecionar uma ou mais análises e, em seguida, clicar no

botão  Edit assignment na parte inferior da página.

Uma janela "Assignment" permite atribuir um ou mais utilizadores às análises selecionadas, procurando o identificador do utilizador (username) através da barra de pesquisa dedicada. Também é possível modificar a atribuição, acrescentando ou eliminando utilizadores. Os utilizadores a quem é atribuída uma análise são listados na coluna "Assigned to". Os utilizadores que nunca iniciaram sessão no SeqOne não podem ser atribuídos até criarem as suas credenciais.

Os utilizadores podem aceder às análises que lhes foram atribuídas (separador To do) e às análises validadas (separador Done) na sua "To do list", acessível através do ícone  no canto superior direito da página. É apresentada uma notificação para cada nova atribuição.

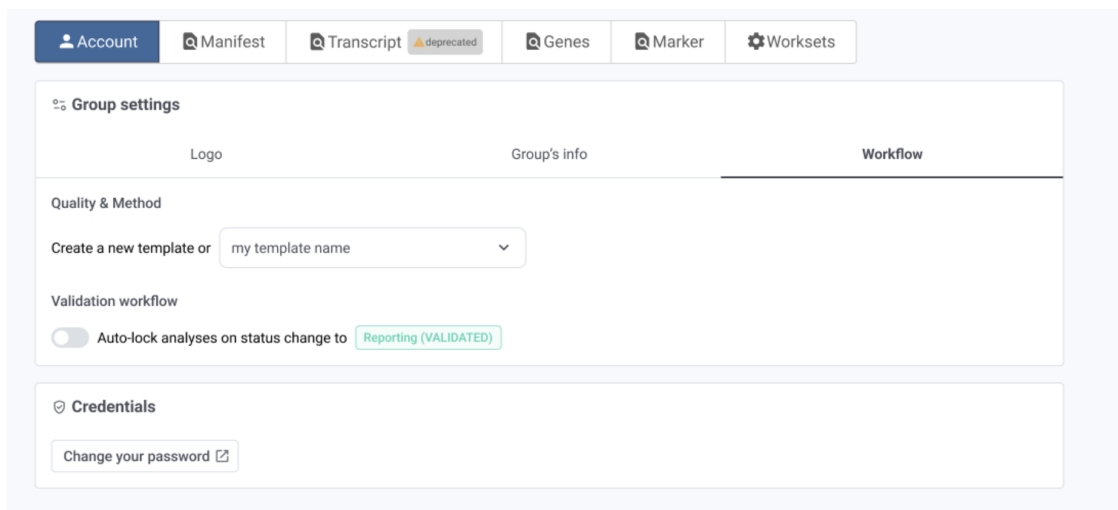
Os utilizadores podem aceder diretamente às análises atribuídas clicando nelas. Podem validar a sua tarefa clicando na caixa de verificação correspondente no separador "To do" para a mover para o separador "Done". Esta ação é reversível.

To do list		
To do 4		Done
	Assigned 10 days ago	☐
	somaHRD - sample_pos_with_BRCA TEST_HRD_sWGS_UMI_xths2	
	Assigned 7 days ago	☐
	somaHRD - sample_pos_with_BRCA TEST_HRD_sWGS_UMI_xths2	

As análises são apresentadas pela seguinte ordem, de acordo com o seu status: concluídas, em curso ou falhadas. Para cada análise, a data de atribuição, o nome da análise e o nome do projeto são detalhados.

13.1.1.2. Bloqueio de análises

O administrador de uma entidade pode bloquear ou desbloquear uma análise. Nas "Settings" da entidade, no separador "Account" e depois em "Workflow", pode ativar o bloqueio automático das análises validadas utilizando a funcionalidade "Auto-lock analyses on status change to Reporting (VALIDATED)":



The screenshot shows the 'Group settings' page with the 'Workflow' tab selected. Under 'Validation workflow', the 'Auto-lock analyses on status change to' toggle is turned on, and the status is set to 'Reporting (VALIDATED)'. Below this, the 'Credentials' section contains a 'Change your password' button.

No final do seu trabalho de interpretação, quando um utilizador altera o estado da análise em "Reporting - validated", a análise é automaticamente bloqueada. As análises bloqueadas apresentam um estado "Reporting - Validated & Locked", visível com o ícone .

 Reporting VALIDATED & LOCKED

Quando uma análise está bloqueada, os utilizadores não podem realizar as seguintes acções a partir da interface :

- adicionar ou modificar avaliações de variantes no VKB',
- acrescentar ou modificar os comentários associados às variantes,
- gerar um novo relatório a partir da janela *Selection*,
- gerar um novo relatório HRD,
- modificar manualmente o estado da análise,
- executar uma das seguintes acções no painel de análise:
 - modificar a atribuição de ensaios (em "Assigned To")
 - modificar o género do paciente
 - adicionar ou modificar HPOs
 - modificar os painéis de análise
 - eliminar relatório clínico

Se um utilizador tentar qualquer uma destas acções, ser-lhe-á pedido que desbloqueie a análise, se tiver permissão, ou que contacte o seu administrador, se não tiver. Um utilizador que pretenda desbloquear uma análise deve ter a permissão "Unlock analyses".

13.1.1.3. Painel lateral para gestão da análise

O painel lateral de análise, também conhecido como "Analysis Drawer", pode ser acedido a partir de qualquer separador de análise.



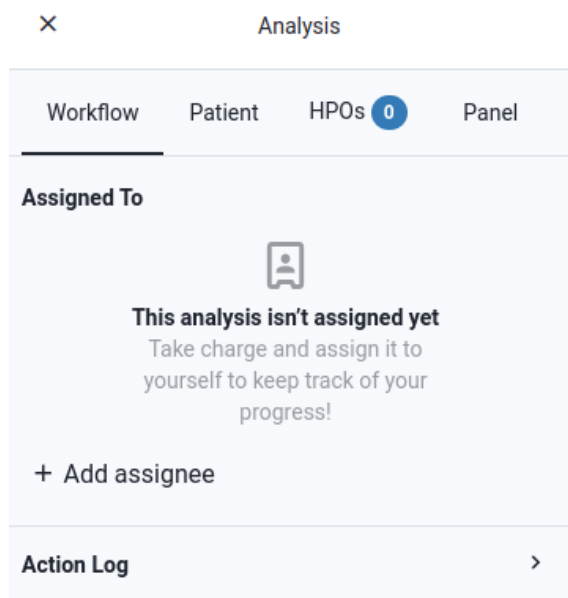
Ao clicar no botão  **Analysis** no canto superior direito do banner dos separadores de análise, o utilizador pode visualizar e gerir diretamente os dados de análise e os 'Patient HPOs' sem sair do ecrã atual.

A gaveta "Analysis Drawer" contém 4 vistas:

- **Workflow:** permite-lhe organizar e rastrear o seu fluxo de trabalho para análise
- **Patient:** permite-lhe ver e modificar os metadados do paciente
- **HPOs:** permite adicionar ou modificar termos HPO na análise, apenas disponível para análises em contexto constitucional.
- **Panel:** permite-lhe aplicar uma lista de genes à análise e filtrar os resultados da análise com base nessa lista.

Visualização modular do Workflow

Esta vista permite-lhe organizar e seguir os fluxos de trabalho dos utilizadores para a análise que está a ser interpretada.



"Assigned To". Esta vista apresenta as tarefas atribuídas para a análise atual. Apenas os administradores ou utilizadores com permissão "Manage assignments" podem atribuir tarefas (ver **Gestão de administradores** - secção de **apresentação da interface**).

Quando as tarefas são atribuídas à análise atual, é apresentada uma lista de todos os utilizadores atribuídos, juntamente com a data e hora de conclusão. Esta lista é actualizada automaticamente assim que um utilizador conclui uma tarefa.

Os utilizadores só podem editar as suas próprias tarefas marcando ou desmarcando a caixa correspondente. Ao passar o rato sobre uma tarefa só de leitura, uma dica de ferramenta indica 'You can only complete yours'.

Se nenhuma tarefa tiver sido atribuída, a vista apresenta "This analysis isn't assigned yet".

Se um utilizador não autorizado tentar atribuir uma tarefa clicando no botão "+ Add assignee", a mensagem de erro indica "You must have administrator privileges to assign analysis. Se necessário, contacte o administrador da plataforma para solicitar o acesso adequado."

Action Log. (Registo de acções). O banner "Analysis Action Log (Audit Trail)" apresenta o histórico de todas as acções realizadas no estado HRD na análise SomaHRD, filtros, perfis de filtros e painéis de genes restritos na tabela de variantes, detalhando o tipo de ação, o utilizador, o carimbo de data/hora e uma descrição da ação.

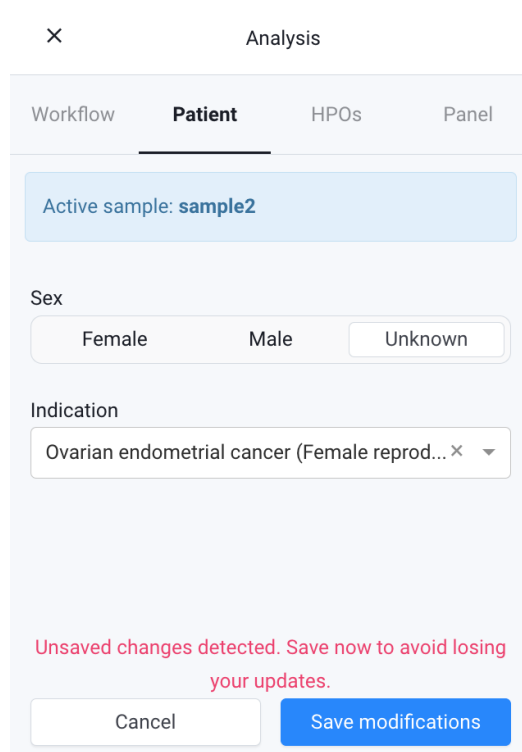
As acções mais recentes aparecem no topo da lista. A paginação permite aos utilizadores navegar eficazmente pelo histórico de acções. O histórico completo pode ser descarregado em formato .csv através do ícone "Download Log".

Visualização modular do Patient

Esta visualização permite visualizar e modificar metadados do paciente, tais como sexo (feminino, masculino ou desconhecido ((Female, Male or Unknown)), o grupo de ascendência genética GnomAD (apenas germinativa) ou a indicação (apenas somática).

A visualização e edição da indicação só está disponível para análises somáticas:

- Se a indicação tiver sido introduzida antes do lançamento da análise (ver "**Sample metadata**"), é definida por defeito para a amostra, é então marcada com uma estrela e aparece na primeira posição da lista.
- Se a indicação for alterada durante a interpretação, uma mensagem de aviso indica que a indicação selecionada é diferente da indicação original e a coloração do comprimido na subcoluna "A" da coluna "*Clinical Evidences*" será actualizada (ver **Selecionar uma indicação**).
- Se não estiver disponível qualquer indicação para análise, é apresentada a mensagem "No available indication".



Analysis

Workflow **Patient** HPOs Panel

Active sample: **sample2**

Sex

Female Male Unknown

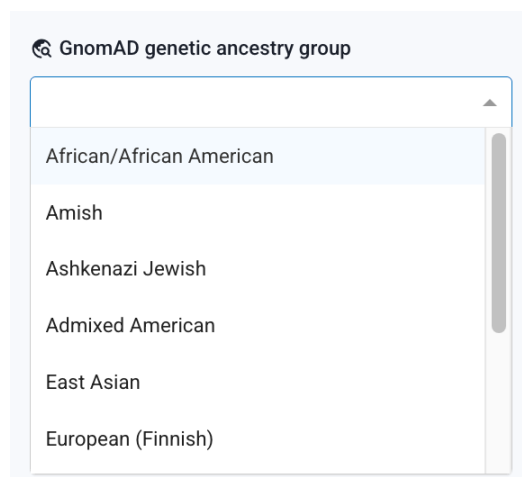
Indication

Ovarian endometrial cancer (Female reprod... ×

Unsaved changes detected. Save now to avoid losing your updates.

Cancel Save modifications

A exibição e edição do «GnomAD genetic ancestry group» (grupo de ascendência genética) só está disponível para análises **germinativas**. O valor padrão está vazio e é tratado como a população GnomAD combinada.



GnomAD genetic ancestry group

African/African American

Amish

Ashkenazi Jewish

Admixed American

East Asian

European (Finnish)

As modificações efectuadas nesta vista devem ser guardadas ou anuladas clicando nos ícones correspondentes "Save modifications/Cancel" na parte inferior do painel, para poder mudar de separador.

Visualização modular do HPOs

Esta vista, disponível apenas em projetos Germline, permite-lhe ver, editar e atribuir termos HPO relacionados com a amostra, manualmente ou a partir de uma descrição clínica do Patient.

O utilizador pode visualizar a lista de termos HPOs já associados à amostra antes da análise. Na ausência de códigos HPOs, é apresentada a mensagem "No HPOs yet".

Para adicionar termos manualmente, o utilizador pode introduzi-los a partir da barra de pesquisa "Enter phenotypes one by one", procurando um fenótipo ou um código HPO. Pode também importar um ficheiro de texto que contenha códigos HPO, clicando no ícone  (ver também a secção '**Sample metadata**').

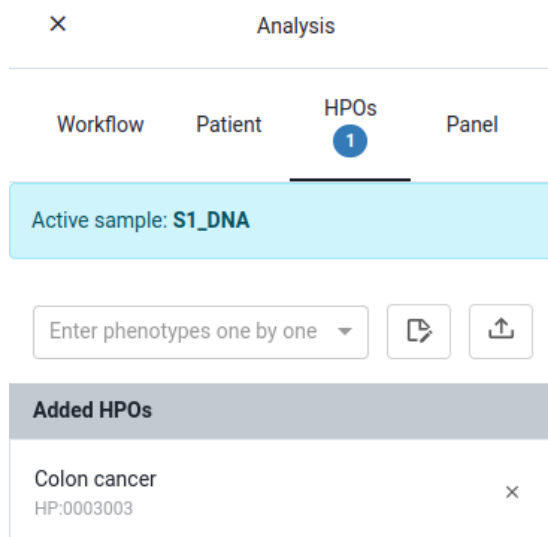


Figura 8: Visão dos HPOs do painel de análise

O utilizador pode gerar termos de HPO com base na descrição clínica do paciente utilizando o DiagAI. Ao clicar no ícone , aparece uma janela de diálogo que pede ao utilizador para descrever o perfil clínico do paciente (esta descrição está limitada a 4000 caracteres) e, em seguida, clicar em "Extract HPOs with DiagAI" para gerar uma lista dos HPOs encontrados. Os termos HPO são destacados a cores e o utilizador pode associá-los ao doente passando o rato por cima deles e clicando em "Add" (Adicionar), ou adicionar todos os termos HPO encontrados clicando em "Add all" (ver também a secção **Editar por metadados de amostras com o painel de análise**).

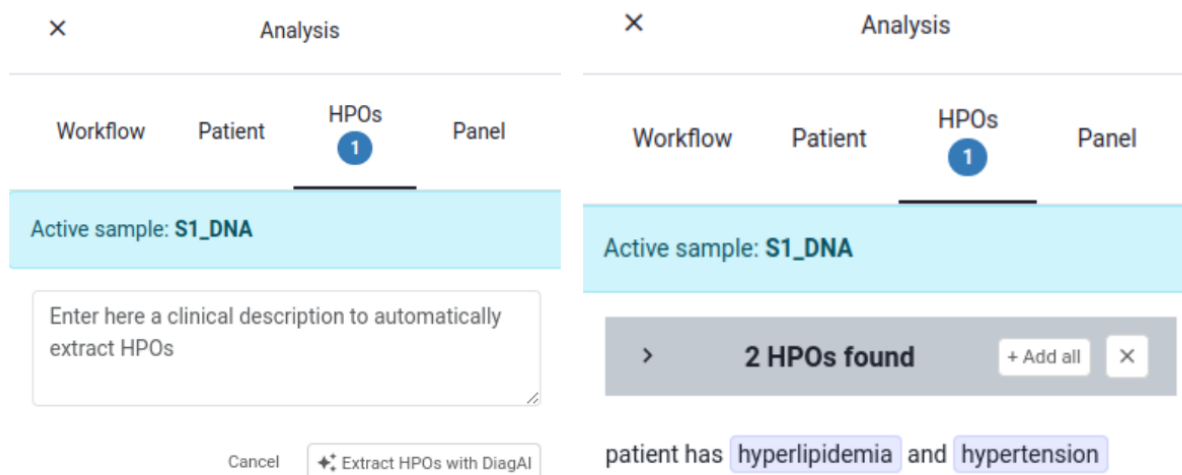
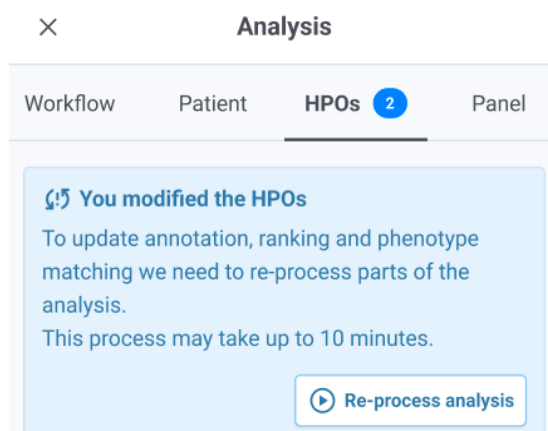


Figura 9: 'Extract HPOs terms with DiagAI' from the HPOs view in the Analysis Drawer".

Quando os HPOs são adicionados ou modificados a partir desta vista, aparece uma faixa azul, sugerindo que a análise seja reprocessada com os novos HPOs.

Ao selecionar "Re-process analysis", o utilizador pode atualizar a pontuação DiagAI e a classificação das variantes (Short list ou SmartPick) de acordo com os novos HPOs introduzidos.

Esta etapa de reprocessamento congela a tabela de variantes durante vários minutos para atualizar a classificação das variantes e a coluna DiagAI".

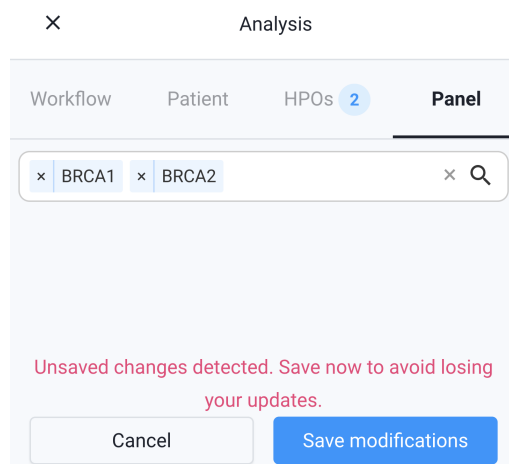


Visualização modular do Panel

Esta vista está disponível para todos os worksets, exceto os worksets de análise de CNV (CNVExome, CNVCapture e SomaCNVCapture).

Dá acesso à funcionalidade "*Restricted panels*", que permite restringir e bloquear a análise a genes selecionados, à maneira de um "*in silico panel*".

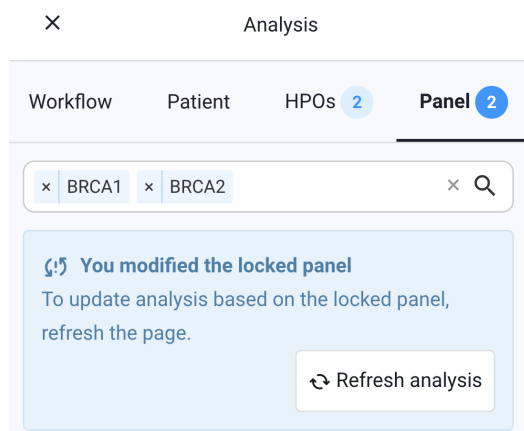
Para isso, o utilizador pode seleccionar, através de um menu pendente, uma ou várias listas de genes a aplicar, entre os importados anteriormente (ver parágrafo **Genes**, na secção **Gestão dos parâmetros do utilizador**). Ao guardar os painéis seleccionados através de "Save modifications", é aplicado um filtro a toda a análise de acordo com o(s) painel(is) de genes seleccionado(s).



Os painéis restritos podem ser seleccionados assim que a análise é lançada (ver secção **Lançamento de uma nova análise**) ou a partir da vista Painel após a análise.

Quando os painéis de genes são adicionados ou modificados, um banner azul avisa o utilizador da alteração ("You modified the locked panel") e sugere a atualização da análise com os novos painéis através da função "*Refresh analysis*".

Ao clicar em "*Refresh analysis*", o utilizador opta por atualizar a tabela de variantes, o separador CNVs" (se aplicável) e o separador GeneCov" de acordo com o novo painel.



13.1.1.4. Pannel lateral para gerir as avaliações

O painel  **Selection** pode ser acedido a partir de qualquer separador dentro de uma análise. Permite ao utilizador visualizar as pequenas variantes e CNVs anotadas no VKB para a análise atual, gerar um ficheiro JSON («Generate file») e editar um relatório clínico («Generate report»).

Os CNVs validados no VKB' também estão disponíveis nesta secção, apenas no caso de CNVs provenientes de análises do exoma, do genoma e de CGH MicroArrays.

As pequenas variantes validadas no VKB" são diferenciadas pelo logótipo  , enquanto as "CNVs" são apresentadas com o logótipo  e as palavras 'gain' ou 'loss'.

× **Selection** 3

[↓ Generate file](#) [☑ Generate report](#)

Variants

 SLC25A22
 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
 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
 103.25kb | SNX14, SYNCRIP

13.1.2. Visão geral dos resultados da análise no separador Dashboard

Para alguns worksets, um separador "Dashboard" que resume todos os resultados e métricas de "QC" está acessível aos utilizadores e aparece por predefinição ao abrir uma análise. O separador "Dashboard" fornece uma visão geral dos resultados da análise em várias secções. As secções disponíveis variam consoante o tipo de workset. Este separador está disponível para SomaCGP, SomaVar Tertiary e SomaLBx.

Patient. Esta secção resume os metadados da amostra: nome da amostra (*Sample ID*), sexo e indicação da patologia (Indication).

Quality Controls. Esta secção fornece informações sobre vários indicadores-chave relacionados com a qualidade da amostra e da sequenciação:

- Pureza do tumor,
- Sample ploidy",
- a percentagem de bases cuja qualidade é superior ou igual a 30 (% bases Q30),
- cobertura média das regiões de interesse (cobertura média ROI) antes e depois do processamento UMI,
- percentagem de regiões de interesse com cobertura superior a 200X (ROI > 200X) antes e depois do processamento UMI.

Signatures. A secção sobre assinaturas apresenta os índices para as análises TMB, MSI e HRD:

- Cálculo do TMB (Tumor Mutational Burden): número de mutações/Mb.
- Microsatellite instability: percentagem de sítios instáveis.
- HRD status (Homologous Recombination Deficiency).

Os estados das pontuações TMB e MSI podem ser selecionados pelo utilizador através dos menus pendentes localizados por baixo de cada pontuação.

Selection. Esta secção apresenta as variantes avaliadas na base de dados VKB para a análise atual.

Suggested by SeqOne. As alterações genómicas detetadas na análise são apresentadas como uma lista reduzida de variantes clinicamente relevantes. As variantes são selecionadas se a sua classe ComPerMed for 5, ou se tiverem uma asserção de Nível I correspondente à indicação do paciente, ou uma asserção de Nível II para terapias experimentais relacionadas com esta indicação, ou uma asserção de Nível I para outra indicação. O utilizador pode adicionar variantes desta categoria à secção Selection (Seleção) clicando em "Add" (Adicionar).

O ficheiro JSON pode ser gerado clicando em "Generate JSON file" e contém todas as informações da página "Dashboard".

13.1.3. Cobertura pormenorizada por exão.

No separador *GeneCov*, os utilizadores podem ver a taxa de cobertura dos genes incluídos no manifesto de análise para cada exão, quer se trate de um painel de genes ou de um exoma.

Metodologia

Este módulo intersecta o manifesto do utilizador com as anotações RefSeq para todos os genes incluídos no painel ou no exoma, mantendo 10 pb de cada lado para cobrir as junções de emenda, e depois calcula a profundidade de cobertura para cada base incluída nesta intersecção utilizando a ferramenta *mosdepth*.

São definidos diferentes limiares para atribuir a cada exão um status (coberto, aviso, falhado) de acordo com a percentagem de bases abaixo destes limiares de cobertura.

O status é definido a partir da cobertura de bases no exão incluindo os 10bp adjacentes. Se for utilizado o UMI, a cobertura é calculada após a dedução.

Os limiares aplicáveis aos diferentes tipos de análises são apresentados em pormenor no quadro seguinte*:

	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%				

Quadro 3: Limiares de cobertura para o módulo *Gene Coverage* de acordo com o tipo de análise.

Se o utilizador tiver predefinido métricas de QC personalizadas a partir dos parâmetros do projeto, estas serão aplicadas no separador *GeneCov*" (ver secção **Configurar métricas de QC predefinidas do projeto).*

Visualização e interpretação

Cada gene pode ser visualizado numa representação gráfica, e os resultados do Gene Coverage podem ser filtrados com base em genes, "in silico panel" e listas de transcritos. Os exões visados são representados a verde (*Covered*), "orange" (*Warning*) ou "red" (*Failed*) de acordo com os limiares descritos acima.

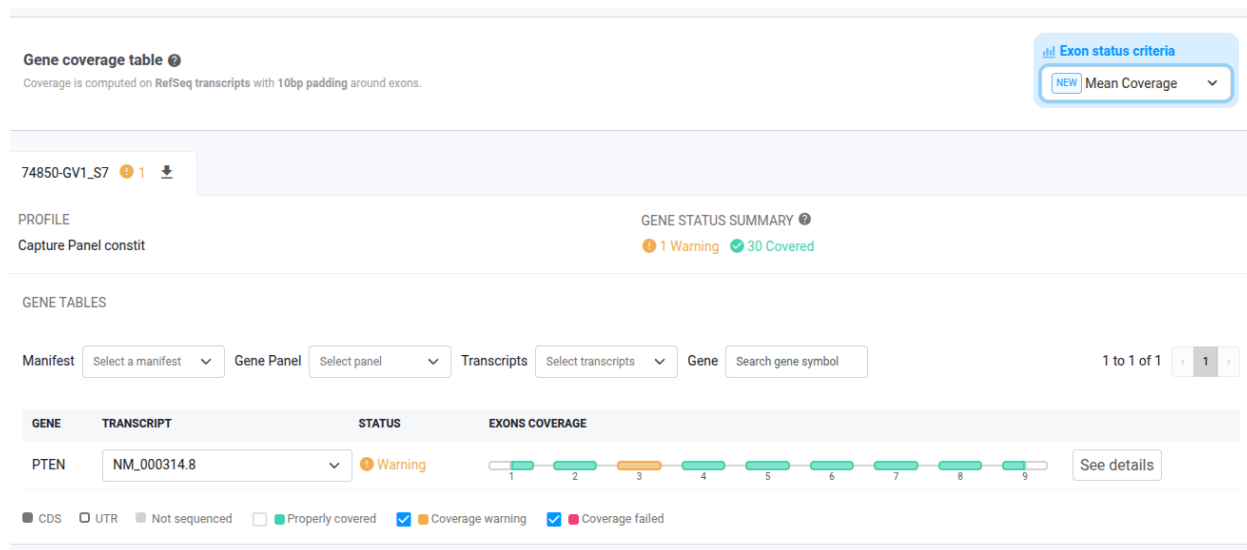


Figura 10: Detalhes do separador GeneCov

Opções de filtro

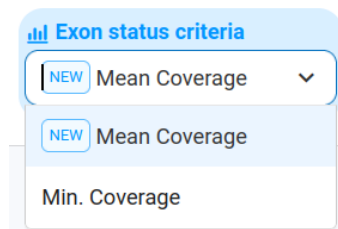
Um menu de filtros permite-lhe:

- restringir a visualização a um subconjunto de regiões utilizando um ficheiro .bed previamente importado para a interface, através do filtro "Manifest",
- aplicar uma lista de genes através do filtro "Gene Panel",
- aplicar uma lista de transcrições através do filtro "Transcripts",
- procurar um gene específico através do filtro "Gene".

Métodos de cálculo da cobertura

O utilizador pode seleccionar o método de cálculo da cobertura a aplicar utilizando o menu pendente "Exon status criteria" no canto superior direito da página:

1. **'Mean Coverage'** (padrão): Utiliza a profundidade de cobertura média em todo o exão para determinar o status da cobertura.
2. **'Minimum Coverage'**: Utiliza a fração mínima de bases que satisfazem um limiar de cobertura específico. Isto garante que uma determinada percentagem de bases tem pelo menos a profundidade de cobertura mínima exigida.



O menu suspenso "**Gene status criteria**" permite aos utilizadores aplicar estes métodos de cálculo ao nível do gene:

- **Mean coverage (default):** Utiliza a profundidade de cobertura média em todos os exões do gene para determinar o estado de cobertura.
- **Worst exon status:** Determina o estado de cobertura do gene com base no exão com pior desempenho. Se algum exão não atingir o limite de cobertura, o gene é considerado não coberto. Este era o único método suportado em versões anteriores.

Dados apresentados

Os exões que não são ou são parcialmente cobertos pelo desenho (< 80% de sobreposição entre o intervalo de manifesto e os limites do RefSeq do exão) são considerados não visados. Estes exões estão representados a cinzento (*Not sequenced*) no gráfico.

Uma vista detalhada de cada gene pode ser utilizada para apresentar as métricas de cobertura de exões clicando em "*See details*": cobertura média, mediana, percentagem de cada região coberta a 10x, 30x, 50x, 100x, 200x, 300x e 500x.

Para cada exão, uma janela IGV' permite aos utilizadores visualizar a cobertura e o alinhamento da sequência com maior detalhe.

Exportação de dados GeneCov

Os dados completos da análise de cobertura estão disponíveis como um ficheiro `sample_genecov.v2.tsv.gz`, acessível a partir do separador Files tab da análise.

A partir do separador GeneCov, o utilizador pode exportar a análise de cobertura de acordo com o método de cálculo da cobertura (Mean Coverage ou Min Coverage), através do ícone



junto ao nome da amostra no canto superior esquerdo do separador da amostra.

O utilizador pode optar por exportar dados de acordo com o estado de cobertura do gene (Failed, Warning, Covered) e exportar apenas a transcrição canónica (Canonical) ou todas as transcrições disponíveis (All) antes de clicar em Download para obter o ficheiro desejado.

Os dados são exportados como um ficheiro .tsv que contém as seguintes colunas:

Download genes from S1_DNA as .tsv file



Genes status

☒ Failed ☒ Warnings ☐ Covered

Transcripts

☒ Canonical ☐ All

Close

Download

- Gene (gene_symbol),
- Número do gene de acordo com o HGNC (gene_id),
- Transcrição canónica (is_canonical_transcript),
- Estado da transcrição (transcript_status)
- exão (exon_number)
- Tipo de exão CDS ou UTR (exon_type)
- Estado do exão coberto ou não coberto (exon_status)
- Região abrangida ou não pelo manifesto (objetivo)
- Posição genómica (interval_id)
- Número de cromossomas (chr)
- Posição inicial (início)
- Posição final (fim)
- Cadeia de caracteres 'forward-reverse' ou 'reverse' (cadeia de caracteres)
- % de cobertura a 1, 10, 20, 30, 50, 100, 200, 300, 500 e 1000X (cov_)
- Cobertura média (cov_avg)
- Cobertura mediana (cov_med)

13.1.4. Variantes pequenas

Metodologia

A detecção de pequenas variantes é efectuada por uma combinação de várias ferramentas, que podem variar de um workset para outro.

Quer se trate de uma análise do exoma ou de um painel de genes, a identificação de variantes é determinada pelo manifesto (em formato BED) associado ao projeto, incluindo os intervalos das regiões de interesse e o preenchimento aplicado a estas regiões (ver secção **Criar um projeto**).

Além disso, as variantes são chamadas tendo em conta a qualidade do alinhamento das reads no genoma humano (versão GRCh37 ou GRCh38). As chamadas "MAPQ0", leituras com uma qualidade de alinhamento (mapeamento) igual a 0, porque se alinham em vários sítios do genoma, não são tidas em conta na identificação de variantes para worksets de análise somática. Nas análises "Germline", as reads MAPQ0 são consideradas para a detecção de variantes se for introduzido um manifesto especificando regiões de alinhamento de baixa qualidade quando o projeto é criado nos metadados Low-Quality Mapping Region. As variantes detetadas a partir de reads MAPQ0 são identificadas na tabela de variantes por uma dica de ferramenta "_mapq0" visível na coluna CAL (Caller).

SNVs / indels. Para a detecção de SNVs e indels :

- *Freebayes*¹ é utilizado como o único variant caller pelos worksets GermlineVar (exome), GermlineFamily (exome), SomaRNA e SomaVar LF. Os limiares de detecção mais baixos são os seguintes
 - VAF $\geq$ 10% em GermlineVar - Exome para o exoma,
 - VAF $\geq$ 5% em GermlineVar para painéis de genes,
 - VAF $\geq$ 1% em SomaRNA,
 - VAF $\geq$ 0,1% em SomaVar LF,
- É acompanhado pelo *GATK HaplotypeCaller*² no workset GermlineVar - Panel.
- O seu homólogo somático, *Mutect2*³, é utilizado juntamente com o Freebayes nos worksets SomaVar, SomaHemato e SomaCGP. O limiar mínimo de detecção é VAF $\geq$ 1% para estes dois chamadores em SomaVar.
- *Sentieon*⁴ é utilizado como um dos principais variant caller pelos worksets GermVar, GermVar Family e SomaLBx.
- *Clair3*⁵ é utilizado como o único variant caller no 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>

As análises GermVar e GermVar Family contam com componentes de bioinformática Sentieon para as principais etapas de análise secundária. Em particular:

- O alinhamento é realizado usando a implementação otimizada do BWA-MEM2 da Sentieon, e as leituras duplicadas são marcadas.
- A chamada germinativa (SNVs/Indels) é realizada usando Sentieon DNAScope⁶.
- O Sentieon TNScope⁷ pode ser usado para detectar variantes em mosaico de baixo VAF (dependendo da configuração do GermVar).

Nota sobre regiões de baixa mapeabilidade (MAPQ0): em regiões altamente homólogas (por exemplo, SMN1/SMN2), as variantes podem ser filtradas pela filtragem de aprendizado de máquina do chamador. Quando necessário, o SeqOne fornece um módulo MAPQ0 opcional (módulo de regiões de baixa qualidade) baseado no Freebayes com limites adaptados para aumentar a sensibilidade em regiões definidas pelo usuário.

As análises SomaLBx contam com componentes de bioinformática Sentieon para as principais etapas de análise secundária de biópsia líquida. Em particular:

- O alinhamento de leitura é realizado com o Sentieon (BWA-MEM), e o pipeline inclui processamento/consenso UMI baseado no Sentieon para suportar fluxos de trabalho de ctDNA de alta sensibilidade.
- A chamada somática de variantes curtas (SNVs/Indels) é realizada usando o Sentieon TNScope⁷ no modo apenas tumor (tumor-only), com suporte de fração alélica muito baixa (VAF $\geq 0,1\%$ dependendo da configuração).
- Nota: os genótipos de variantes são relatados como heterozigotos por padrão nos resultados SNV/indel do SomaLBx porque o Sentieon TNScope não realiza previsão de genótipo.

SVs - Structural Variants. As inserções e as deleções cujo comprimento é próximo ou superior ao de uma leitura de sequenciação são detetadas utilizando o SV caller GRIDSS⁸ nos worksets GermlineVar (painel/exoma), SomaVar, SomaVar LF SomaHemato e SomaCGP. Esses eventos são detetados usando o SV caller Delly⁹ nos worksets GermVar e GermVar Family. Apenas inserções e deleções menores que 300bp são detetadas em worksets somáticos, enquanto um limite de 10kb é aplicado no worksets GermlineVar (versões ≥ 2.0), e worksets GermVar e GermVar Family (versões ≥ 3.5). As variantes detetadas pelo GRIDSS são detetadas independentemente do VAF.

FLT3-ITD. Um módulo de detecção específico para duplicações em tandem internas do gene FLT3 (FLT3-ITD) está disponível apenas nos worksets SomaVar e SomaHemato, a partir da versão v2.2. Este módulo baseia-se na ferramenta *FiLT3r*¹⁰.

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>

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

Alu Insertions. Por último, foi também desenvolvido pela SeqOne um módulo de detecção de inserções de elementos transponíveis (*AluMEI*'), como os elementos Alu. É utilizado pelos worksets GermlineVar (painel/exoma), GermlineFamily (painel/exoma), GermVar (painel/exoma/genoma), GermVar Family (painel/exoma/genoma), SomaHemato e SomaVar. Não é aplicado um limiar inferior de VAF às variantes detetadas pelo *AluMEI*'.

As variantes combinadas do *Freebayes*, *HaplotypeCaller* (ou *Mutect2*) e *FiLT3r* caller estão disponíveis no ficheiro *final.vcf.gz* da análise. As variantes específicas do *GRIDSS*, *FiLT3r* ou *AluMEI* estão disponíveis nos respectivos ficheiros VCF (**gridss.vcf.gz**, **filt3r.vcf.gz** e **variants.alumei.vcf.gz**), também disponíveis para download no 'Files tab' da análise.

No caso específico do '*FiLT3r*' :

- As variantes *FiLT3r*' comuns a *Freebayes* e *Mutect2*' estão disponíveis no VCF final da análise (**final.vcf.gz**),
- As variantes exclusivas do *FiLT3r*, bem como as variantes comuns ao *FiLT3r* e ao *GRIDSS*, estão listadas no VCF do *FiLT3r* (**filt3r.vcf.gz**).

Dois módulos adicionais desenvolvidos pela SeqOne permitem então :

- Retenção de variantes multinucleotídicas, ou MNVs, que são codificantes e restritas a um único códon durante o processo de decomposição de variantes do tipo delin.
- Recalibração da frequência alélica das variantes indels e delins detetadas pelo *Freebayes* e pelo *GATK*' (ferramenta *Kraf*).

Estes dois módulos dizem respeito apenas aos worksets GermlineVar (painel/exoma), SomaVar, SomaRNA e SomaVar LF.

O *Kraf* utiliza uma estratégia de k-mers únicos, sem alinhamento. A frequência alélica recalibrada de indels e delins só é retida se for superior à frequência alélica inicial da variante. As variantes são depois anotadas principalmente utilizando a ferramenta *Variant Effect Predictor*¹¹ (VEP), sendo acrescentadas informações adicionais das seguintes bases de dados:

- ClinVar
- OMIM (versão comercial) + PanelApp Australia + PanelApp England
- CKB
- dbSNP
- COSMIC (versão comercial)
- GnomAD" (dados do exoma + do genoma)
- Dfam
- DGV
- LOVD

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

Além disso, estão disponíveis as seguintes pontuações *in silico*:

- Pontuações de splicing (*MaxEntScan*, *GeneSplicer*, *dbSCSNV*)
- Pontuação de conservação (*GERP++*, *PhastCons100way*, *PhastCons20way*, *PhyloP100way Vertebrate*, *PhyloP20way Mammalian*, *SiPhy*)
- Previsões funcionais (*REVEL*, *DANN*, *FATHMM*, *FATHMM-MKL*, *LRT*, *MetaLR*, *MetaSVM*, *MutationAssessor*, *MutationTaster 2*, *PROVEAN*, *SIFT*, *PLI*)

Visualização e interpretação

O separador Variants  permite-lhe visualizar, filtrar e anotar as variantes obtidas durante uma análise. Estas são apresentadas sob a forma de tabela, correspondendo cada linha a uma variante diferente. O número de variantes selecionadas após a filtragem (variantes selecionadas) e o número total de variantes detetadas (número total de variantes) são apresentados no canto superior direito do cabeçalho da tabela de variantes utilizando o seguinte ícone: . Uma paginação na parte inferior da página permite-lhe navegar de uma página de resultados para a seguinte. Se a análise utilizar um painel restrito, o cabeçalho da tabela também incluirá o número total de variantes no painel restrito (*total number of variant in the restricted panel*) .

Esta navegação é facilitada por atalhos de teclado utilizando as teclas ← e → do teclado.

Por defeito, são apresentadas várias colunas descritivas na tabela. Os utilizadores podem exportar a tabela como um ficheiro TSV, clicando em .

Podem estar disponíveis vários transcripts para cada variante e o seu número é apresentado à esquerda de cada linha. As variantes são anotadas nos transcripts RefSeq, com exceção das variantes mitocondriais que são anotadas no *Ensembl*.

Os transcripts do MANE Select e do MANE Clinical Plus são

identificados por um ícone,  ou , respetivamente, acompanhado de uma dica de ferramenta.

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

Para a mesma variante, exceto para as variantes mitocondriais, os transcripts são ordenados de acordo com a ordem abaixo:

- o impacto ['high' > 'moderate' > 'low' > 'modifier']
- type of transcripts ["NM", "XM", "NR", "XR"] (para transcripts RefSeq)
- Tipo de transcript MANE [Mane Select > MANE Clinical Plus > non MANE] 'Mane Select > MANE Clinical Plus > non MANE'
- ordem alfanumérica

Para as variantes mitocondriais, os transcripts são ordenados pela seguinte ordem:

- impacto ['high' > 'moderate' > 'low' > 'modifier']
- se o impacto for 'modifier':
 - o efeito "non_coding_transcript_exon_variant".
 - o biótipo "protein coding > Mt_rRNA > Mt_tRNA"

c) ordem alfanumérica

DiagAI. Esta coluna mostra a pontuação de 0 a 100 com base no conjunto DiagAI que utiliza a aprendizagem automática. A pontuação DiagAI baseia-se em quatro componentes:

- O algoritmo UP², que prevê a patogenicidade da variante
- A matriz PhenoGenius, que avalia a correspondência entre o gene e as HPOs do paciente (se estas tiverem sido introduzidas durante a análise)
- A concordância entre o genótipo observado e o modo de transmissão da doença descrito nas bases de dados GenCC, OMIM e PanelApp (Australia e England).
- A qualidade da variante

Quanto mais próxima de 100 for a pontuação, mais se prevê que a variante seja patogénica e de interesse para o diagnóstico. Quanto mais próximo de 0, mais benigna se prevê que a variante seja.

Dois ícones chamam igualmente a atenção para determinadas variantes de interesse:

- O ícone  significa que a variante tem uma maior probabilidade de ser a causa da doença (AI Smart Pick).
- O ícone  indica que a variante está presente na lista curta de variantes com uma elevada probabilidade de ser a causa da doença do paciente.

Os termos HPO podem ser introduzidos antes do início da análise, editando os metadados da amostra (ver a secção **Sample metadata**). Também podem ser editados durante a interpretação utilizando a vista modal HPOs no painel lateral da análise (ver secção **HPOs modal view**). Neste caso, o utilizador pode reprocessar a pontuação DiagAI e a classificação das variantes (Short list ou SmartPick) de acordo com os novos HPOs introduzidos, utilizando a função "Re-process with new HPOs".

Se o utilizador clicar no ícone da pontuação DiagAI, a página da variante abre-se diretamente no separador DiagAI que detalha a pontuação (ver DiagAI na secção da página da variante abaixo).

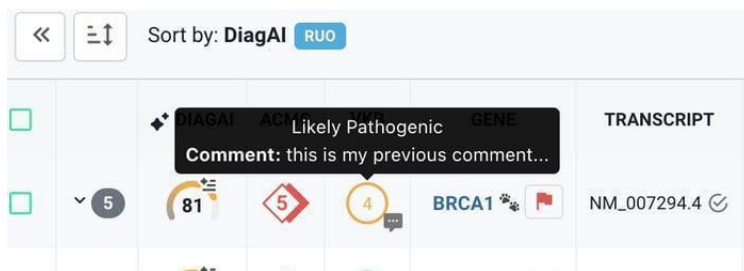
DIAGAI
 86
 81
 78
 78

ACMG. A classificação de patogenicidade de cada variante é calculada como parte de um processo semi-automatizado de acordo com as recomendações emitidas pelo *American College of Medical Genetics and Genomics*. Os utilizadores podem rever a classificação num separador dedicado na página da variante, que é depois utilizado para gerar a avaliação final de uma variante.

ComPerMed. A classificação da patogenicidade de cada variante é calculada como parte de um processo semi-automatizado de acordo com as recomendações emitidas pelo comité belga de peritos ComPerMed (Personalised Medicine Commission). Os utilizadores podem aceder a informações detalhadas sobre a classificação num separador específico na página da variante.

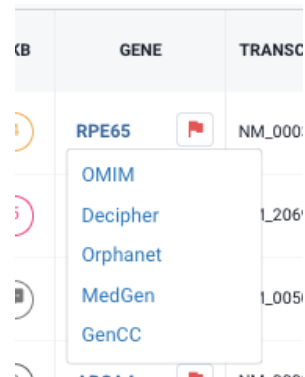
AMP. (Versão de análise SomaVar v1.8 ou posterior, SomaVar LF ou SomaCGP apenas). Esta coluna é apresentada por defeito nas análises que contêm anotações CKB e mostra o nível AMP de cada variante, quando disponível (Nível I ou II). O nível de AMP apresentado baseia-se nas melhores evidências clínicas em todos os tipos, independentemente da indicação selecionada.

VKB. As Evaluations feitas pelo utilizador durante as análises anteriores são armazenadas na base de conhecimentos do utilizador. Essa avaliação aparece na coluna VKB' da tabela para cada nova ocorrência da variante. Uma dica de ferramenta exibe rapidamente a classe e o comentário associados à variante avaliada ou avaliada anteriormente.



Gene. Nomenclatura dos genes de acordo com o [HGNC](#) e sinalização OMIM associada :

- Bandeira vermelha: "OMIM gene" **com** "phenotypes" e "clinical characteristics",
- Bandeira de contorno vermelha: "Gene OMIM" **com** phenotypes mas **sem** características clínicas ",
- Ponto vermelho: "Gene OMIM" **sem** phenotypes ou características clínicas ".



Os genes conhecidos por serem genes de imprinting de acordo com PanelApp (Australia) são destacados pelo ícone . A dica de ferramenta associada indica se o gene é expresso maternalmente, expresso paternalmente, desconhecido ou se são referenciados vários

imprintings.

Clicar na bandeira junto ao nome do gene leva-o diretamente para a página OMIM. Ao clicar no nome do gene, aparece um menu pendente que fornece ligações para as bases de dados OMIM¹², Decipher¹³, Orphanet¹⁴, MedGen¹⁵ e GenCC¹⁶. Cada ligação conduz diretamente à página correspondente ao gene selecionado.

As informações relativas ao OMIM e às bases de dados só estão disponíveis para análises de linha germinal.

HGVSC. Esta coluna fornece a anotação HGVS da variante no ADN complementar (cDNA) e na proteína.

HGVSC

c.10580G>A
p.R3527Q

Effect. Esta coluna descreve o impacto da variante na proteína e o efeito no transcript.

O impacto é calculado de acordo com 4 níveis:

- **High** (high impact, a vermelho): a variante tem um elevado impacto disruptivo na proteína, provavelmente causando o truncamento da proteína, a perda de função ou despoletando o seu decaimento mediado por nonsense.
- **Moderate** (moderate impact, a amarelo): variante não perturbadora suscetível de modificar a eficácia da proteína.
- **Low** ("baixo impacto", a cinzento): variante considerada inofensiva ou pouco suscetível de modificar o comportamento da proteína.
- **Modifier** (impacto modificador, a azul): geralmente variantes não codificantes ou variantes que afectam regiões ou genes não codificantes, para as quais as previsões são difíceis ou não há provas de impacto.

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

Para cada variante, podem estar disponíveis vários efeitos no transcript' na coluna Effect. Neste caso, apenas é apresentado o efeito com a consequência mais elevada de acordo com o PEV¹⁷, seguido de um número (por exemplo, +2) para os outros efeitos disponíveis. Os pormenores

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

destes efeitos são mostrados na dica de ferramenta correspondente.

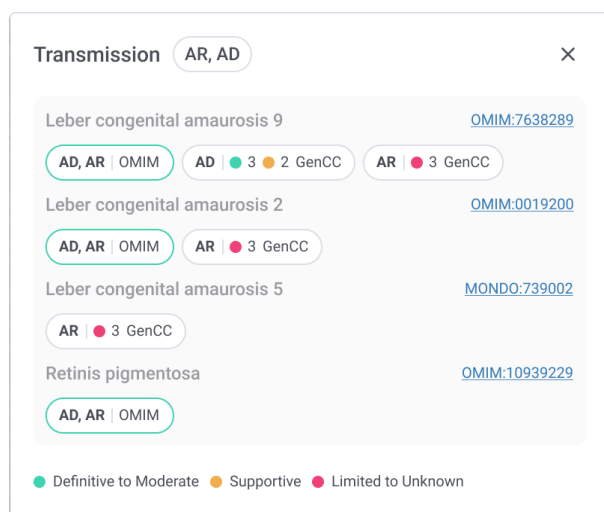
Transmission (Transmissão). Esta coluna indica o modo de transmissão de doenças associadas a genes registrados nas bases de dados GenCC e OMIM. Os emblemas coloridos são visíveis para identificar rapidamente os padrões de transmissão com base nas previsões do DiagAI:

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

Os emblemas De Novo, Compound Het - Verified" e "Recessive Hom" só são apresentados em análises de famílias.

Pode ser utilizado um filtro para selecionar variantes com estes emblemas (ver **seção Filtrar variantes - Transmission & Phenotypes**).

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



Clicando nas ligações dos modos de transmissão nesta coluna, abre-se uma janela pop-up onde os utilizadores podem encontrar doenças que estão listadas com os seus nomes e ligações a bases de dados (OMIM, MONDO - se o OMIM não estiver disponível - ou GenCC).

As doenças estão organizadas numa hierarquia específica: as que se encontram tanto no OMIM como no GenCC aparecem em primeiro lugar, seguidas das entradas apenas

no OMIM e depois das entradas apenas no GenCC.

Cada entrada inclui etiquetas codificadas por cores: As etiquetas OMIM têm contornos verdes, enquanto as etiquetas GenCC utilizam um sistema de semáforo (verde/amarelo/vermelho) para indicar a força das provas:

- Verde: Definitivo + "Strong" + "Moderate"

- Amarelo: "Supporting
- Red: Limitado, Animal, Unknown)

Clinical Evidences. (Análise SomaVar v1.8 ou posterior, SomaVar LF ou SomaCGP apenas). Esta coluna é apresentada por defeito nas análises que contêm anotações CKB. Está dividida em três sub-colunas, cada uma contendo o número total de um determinado tipo de evidência: A (Actionability), D (Diagnosis) e P (Prognosis). A coluna A também apresenta o número de asserções terapêuticas sensíveis (esquerda) e resistentes (direita) como um distintivo.

Os valores nestas subcolunas são codificados por cores:

- A resistência, quando encontrada (> 0 asserções), é sempre apresentada a vermelho, independentemente da indicação, e a cinzento, caso contrário.
- A sensibilidade é realçada a verde apenas quando a indicação correspondente está selecionada, e a cinzento caso contrário.

Se pelo menos uma prova deste tipo corresponder à indicação selecionada na tabela de variantes (painel Indicação), aparece um visto na dica de ferramenta à frente da etiqueta da indicação. Se não for encontrada nenhuma correspondência, a etiqueta não aparece.

Phenotypes. Esta coluna, disponível para projetos Germline, permite a associação de um gene a termos HPO utilizando a matriz PhenoGenius¹⁸. Se os termos HPO estiverem associados à amostra, a percentagem mais elevada de correspondência entre o gene e o fenótipo é apresentada, juntamente com o rácio de termos aos quais o gene está associado. Se nenhum gene corresponder às HPOs introduzidas, a coluna apresenta a mensagem "no match"; se a amostra não tiver termos de HPO, a coluna apresenta a mensagem "see HPO". Em ambos os casos, os termos conhecidos para o gene são, no entanto, propostos na janela pop-up associada.

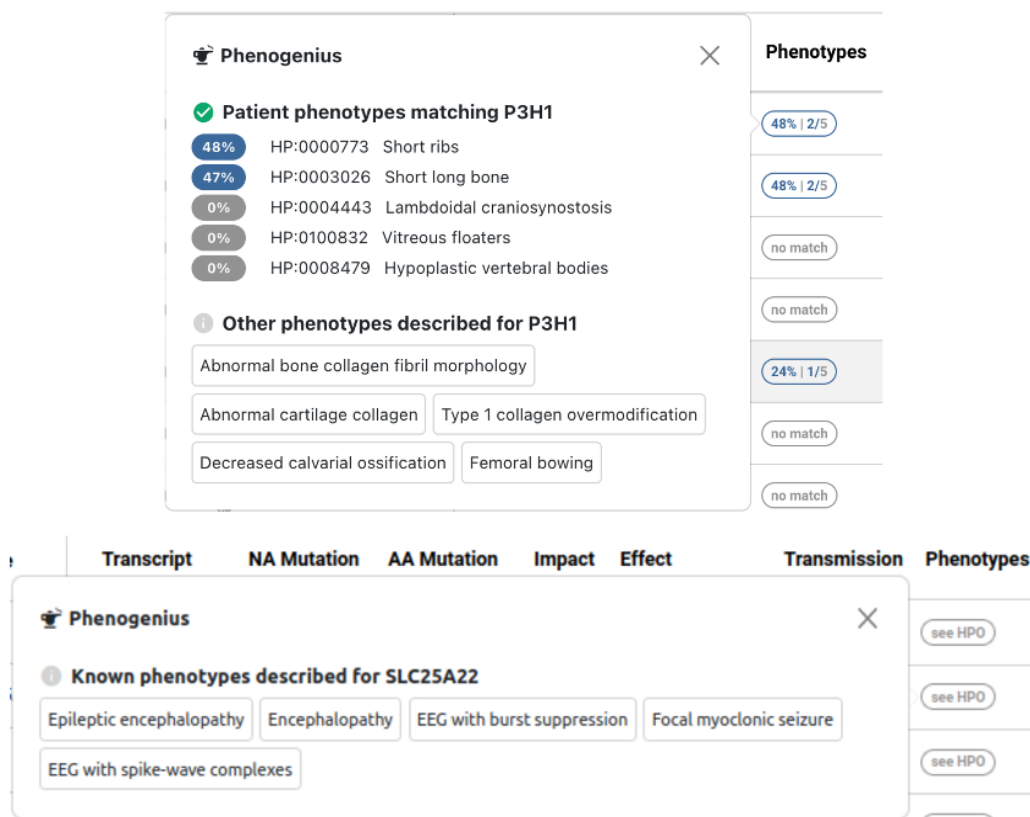
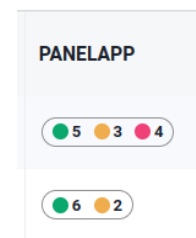


Figura 11 : Visualizar informações sobre o PhenoGenius

¹⁸ Kevin Yaury *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 coluna integra anotações da base de dados de painéis *PanelApp Australia*¹⁹ e *PanelApp England*²⁰. Indica o número de painéis associados ao gene para o qual foi detetada uma variante, de acordo com o nível de classificação descrito abaixo.

Se um gene não for encontrado num painel, a coluna apresentará "N/A".



A janela pop-up associada apresenta detalhes da associação gene/fenótipo, bem como o modo de transmissão associado ao fenótipo, o modo de transmissão associado ao fenótipo e a base de dados da qual o painel é derivado, "PanelApp Australia" (**AUS**) ou "PanelApp England" (**ENG**) (**Figura 12**).

A cada gene será atribuído um ou mais painéis de acordo com um sistema de classificação de três níveis baseado na evidência da associação entre o gene e o fenótipo: Evidências elevadas a **verde**, evidências moderadas a **amarelo** e evidências baixas a **vermelho**.

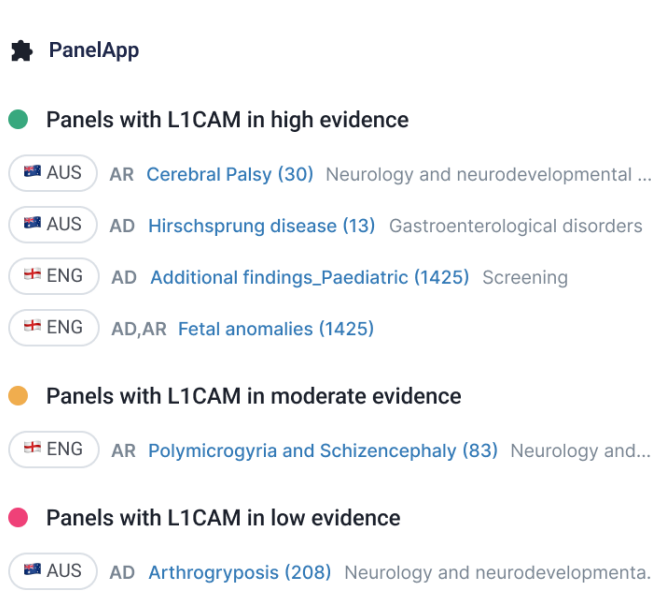
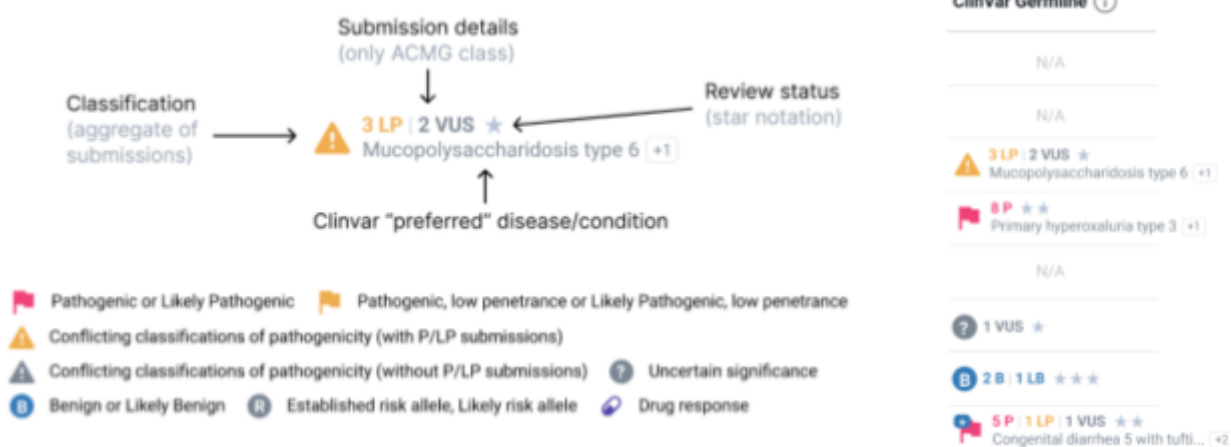


Figura 12: Visualização das informações da PanelApp.

¹⁹ <https://panelapp-aus.org>.

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

ClinVar Germline. Esta coluna apresenta informações extraídas da ClinVar durante a análise. Fornece uma série de informações, incluindo a classe de patogenicidade, os detalhes e o número de submissões para cada classe ACMG, o status da revisão e a condição/doença preferida da ClinVar. Ao clicar na célula da tabela de variantes, o utilizador será redireccionado para a página ClinVar correspondente.



Para mais informações, consultar https://www.ncbi.nlm.nih.gov/clinvar/docs/review_status/

GnomAD. Esta coluna apresenta 6 elementos da base de dados GnomAD para uma variante específica na população: a frequência alélica (AF), o número de alelos (AC), o número de homozigotos (HOM), o número de hemizigotos (HEM), a frequência do alelo majoritário (MAF) e, para os projetos GRCh37, a versão do GnomAD para a qual esta variante foi encontrada (VER).

Ao clicar no valor apresentado na coluna AF, o utilizador é redireccionado para a base de dados GnomAD na sua versão v4 ou, na sua falta, na versão v2. Se uma variante não estiver listada na base de dados, o valor apresentado nesta coluna será 'N/A'.

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

É apresentado um alerta na coluna VER se esta variante estiver presente no GnomAD, mas não passar em todos os filtros de qualidade para pelo menos um dos dois conjuntos de dados (exome ou genoma). Neste caso, os valores apresentados adaptam-se aos valores disponíveis para esta variante e os filtros que accionaram o alerta são apresentados, juntamente com uma breve mensagem que

Menciona o(s) conjunto(s) de dados infractores.

Score UP². A pontuação de previsão funcional UP² (Universal Pathogenicity Prediction), que varia de -1 a 1, classifica a patogenicidade das variantes utilizando uma abordagem de aprendizagem automática de acordo com três categorias : **P** (Pathogenic : pontuação > 0,1), **VUS** (Variant of Unknown Significance : pontuação ≥ -0,1 et ≤ 0,1) ou **B** (Benign : pontuação < -0,1). A pontuação UP² é apresentada por defeito na tabela de variantes, mas o utilizador pode

desativar a coluna UP² a partir do painel de gestão de colunas, clicando no botão



CI-SpliceAI. Esta coluna incorpora pontuações de previsão calculadas através de uma abordagem de aprendizagem automática para identificar variantes com impacto no splicing alternativo à maneira do *CI-SpliceAI*'.²¹

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

Os Scores" são calculados para um subconjunto de variantes que satisfazem os seguintes critérios: variantes de emenda de acordo com VEP (splice_*) com uma frequência GnomAD" inferior a 10⁻³ e que satisfazem os seguintes critérios de qualidade (QA > 20, AO > 10, DP > 20).

Se calculada, a coluna fornece 4 pontuações delta para a variante em causa, para o ganho do aceitador (AG), perda do aceitador (AL), ganho do dador (DG) e perda do dador (DL).

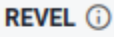
A pontuação delta de uma variante, que varia de 0 a 1, corresponde à probabilidade de a variante modificar o splicing. É aplicado um código de cores aos limiares da pontuação delta: **verde** (>0,2; alta recordação), **amarelo** (>0,5; recomendado) e **vermelho** (>0,8; alta precisão). Se a pontuação delta for inferior a 0,2, consideramos que a pontuação é insignificante. A coluna especifica então 'No splicing impact'.

Para cada variante, é fornecida uma hiperligação para a interface de pesquisa *SpliceAI*'²² correspondente.

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

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

REVEL. A pontuação de previsão REVEL baseia-se num método global de previsão da patogenicidade de variantes missense utilizando uma combinação de pontuações de 13 ferramentas individuais: MutPred, FATHMM v2.3, VEST 3.0, PolyPhen-2, SIFT, PROVEAN, MutationAssessor, MutationTaster, LRT, GERP++, SiPhy, phyloP e phastCons.

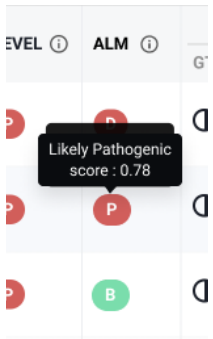


B

P

A pontuação pode variar entre 0 e 1 para uma variante missense, sendo que pontuações mais elevadas reflectem uma maior probabilidade de a variante causar a doença. Os limiares aplicados são os seguintes: REVEL score > 0,5 para **P** (Pathogenic) e REVEL score < 0,5 para: **B** (Benigna). A pontuação pode ser apresentada na tabela de variantes pelo utilizador assinalando a caixa REVEL na inserção "Columns", na secção "Functional Prediction".

ALM. A pontuação de previsão do AlphaMissense²³ está disponível na coluna ALM. O AlphaMissense é um modelo de previsão que avalia a patogenicidade das variantes missense utilizando a aprendizagem profunda. O modelo é construído com base nas previsões estruturais do AlphaFold para melhorar a sua precisão.



O AlphaMissense integra a conservação evolutiva e o contexto estrutural para avaliar o potencial impacto das variantes missense na função da proteína.

A pontuação varia entre 0 e 1, com valores mais elevados a sugerir uma maior probabilidade de uma variante ser deletéria. Os limiares aplicados são: ALM score > 0,564 para **P** (Likely Pathogenic) e ALM score < 0,34 para **B** (Likely Benign). Noutros casos (0,34 < pontuação ALM < 0,564), será rotulado de **VUS** (Ambiguous).

Sample Freq. Esta coluna mostra, para cada variante, a sua frequência no projeto (sub-coluna PROJECT) e em todos os projetos da entidade com metadados semelhantes (sub-coluna "ALL").

A "Sample frequency" do projeto (PROJECT) corresponde ao número de amostras em que a variante é detetada em relação ao número total de amostras analisadas no projeto.

A "Sample frequency" de amostras ALL corresponde à percentagem de

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

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

amostras analisadas em que a variante é encontrada na entidade, dentro de projetos de metadados semelhantes. Ao clicar no valor apresentado na coluna "ALL", o utilizador pode aceder aos nomes dos projetos e ao número de amostras em que a variante está presente, bem como aos detalhes de todos os projetos de metadados semelhantes (opção "Show all projects").

Por predefinição, a subcoluna ALL está oculta na tabela de variantes, mas pode ser reactivada a partir do painel de gestão de colunas, clicando no ícone  e activando o botão "Sample Frequency Entity (Entity)".

Exon. Esta coluna indica o número total de exões no gene de acordo com o transcript em causa, bem como o número afetado pela mutação, sob a forma de uma fração. Uma dica de ferramenta indica a distância entre a variante e o sítio de splicing mais próximo. Se a variante estiver a montante do local de splicing, o valor será positivo e se estiver a jusante, será negativo.

GT (Genotype). Essa coluna mostra o genótipo previsto pelos programas de deteção de variantes. Os vários pictogramas possíveis são descritos na secção **Variantes de filtragem**.

Numa análise em trio (GermlineFamily), se for possível determinar a origem de uma variante heterozigótica, esta será indicada pelo seguinte código de cores:

- **pink** → Origem materna heterozigótica
- **blue** → Origem paterna heterozigótica
- **red** → Variante heterozigótica *de novo*
- **orange** → violação mendeliana

Não é aplicada qualquer cor se a transmissão for indeterminada (por exemplo, se ambos os pais forem heterozigóticos), ou para variantes homozigóticas.

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

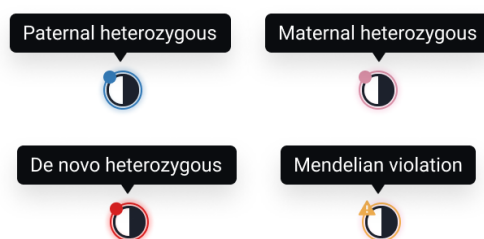


Figura 13: Código de cores aplicado aos genótipos heterozigóticos com base na transmissão GermlineFamily

O faseamento de variantes está disponível para análises GermVar Tertiary e GermVar Tertiary Family utilizando ficheiros VCF dos sequenciadores da Oxford Nanopore Technologies. Quando as variantes são faseadas, os ícones de genótipo aparecem da seguinte forma : . A marca de verificação indica que a variante está em fase. As variantes em fase com a parte preta no mesmo lado e pertencentes à mesma "Phase ID" estão em fase, ou seja, na mesma cadeia.

Pode clicar no genótipo para obter uma lista de variantes cis ou trans, se presentes. Apenas as 5 variantes com a classificação mais elevada do DiagAI são apresentadas, como se mostra ao lado.

A lista completa das variantes está disponível clicando em "Browse all variants".

Variants in trans
Phase set ID 283928 X
GT

! Compound heterozygous candidates for

4 SLC25A22 chr11:g.792142C>T

5 SLC25A22 chr11:g.792146_792147del

VOF (Variant Allele Frequency). Esta coluna indica a frequência do alelo variante, ou seja, a proporção de reads que suportam o alelo alternativo (AO) em relação ao número total de reads nessa posição (COV), expressa em percentagem.

As análises que utilizam os worksets SomaVar versão ≥ 2.5 , SomaCGP versão ≥ 0.7 e SomaVar LF versão ≥ 2.1 exibem as seguintes informações adicionais:

- uma dica de ferramenta apresenta os valores VAF, AO e COV calculados pelos 2 *variant callers* 'Freebayes' e Mutect2' que detectam a variante.
- um ícone de aviso  alerta o utilizador para qualquer desacordo entre os 2 *variant callers* para variantes detetadas com um VAF baixo. Assim, para as variantes que apresentam um VAF inferior a 5% na coluna VAF, o aviso é apresentado no caso de uma discrepância de VAF $> 10\%$ entre os VAFs calculados pelo Freebayes e pelo Mutect2.

AO (Alternate Observation). Esta coluna indica o número de reads que transportam a variante.

CAL (Caller). Esta coluna indica o nome do(s) *chamador(es)* de variantes que detetou(aram) a variante.

Cov (Coverage). Esta coluna indica o número total de "reads" na posição da variante.

SOR. O Strand Odds Ratio (SOR) é uma métrica de viés de cadeia derivada do teste de Odds Ratio Simétrico que mede o desequilíbrio das leituras de suporte à variante entre as cadeias direta (forward) e reversa (reverse). Um SOR mais elevado sugere uma maior probabilidade de um artefacto, enquanto um SOR mais baixo sugere um suporte de cadeia mais equilibrado e é mais consistente com uma variante verdadeira.

Ordenação de variantes

Existem várias opções para ordenar variantes no separador Variantes. Estas estão disponíveis num menu pendente no canto superior direito da tabela. Para cada opção de ordenação, é possível ordenar por ordem crescente ou decrescente clicando no ícone de ordenação . Por defeito, todas as opções são ordenadas por ordem decrescente, com exceção da opção 'Locus', que é ordenada por ordem crescente.

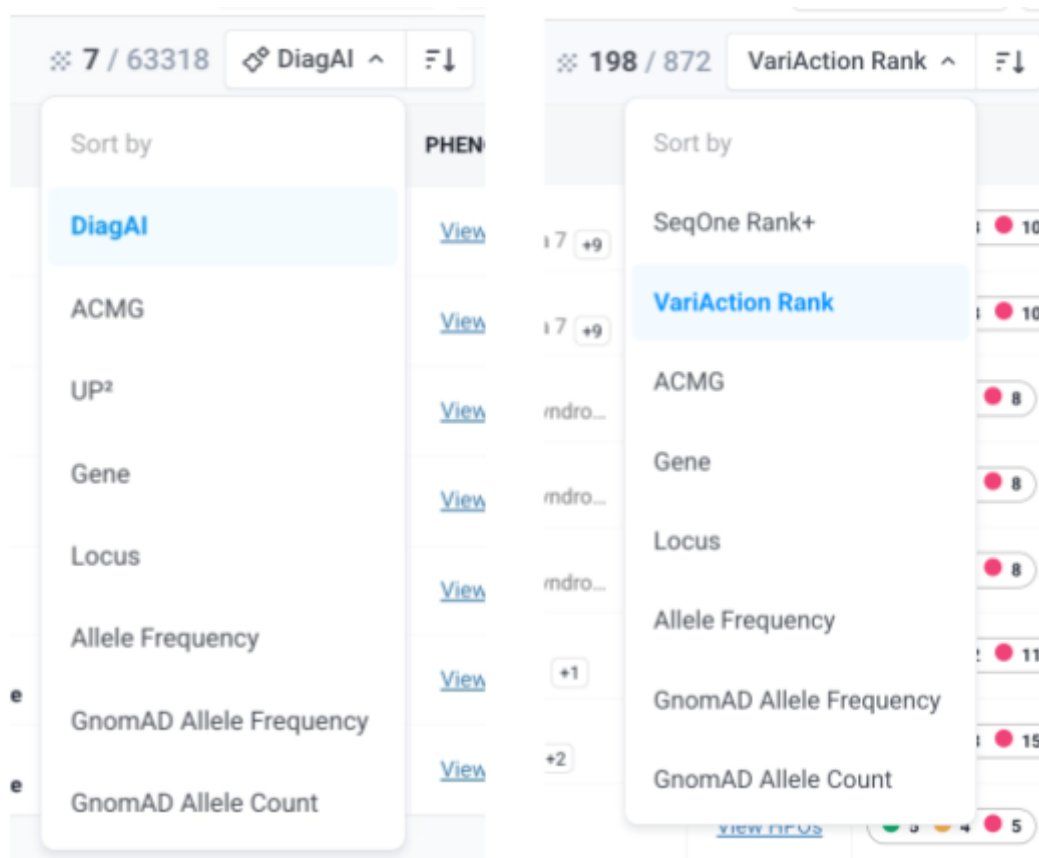


Figura 14: Janela de seleção da opção de ordenação de variantes

DiagAI Score. RUO Esta opção de ordenação é aplicada por defeito a todas as análises de linha germinal e permite que as variantes sejam ordenadas de acordo com a sua pontuação DiagAI (substitui o SeqOne Rank+ nas análises efectuadas após 2 de outubro de 2024).

SeqOne Rank+. RUO O algoritmo de atribuição de prioridades SeqOne Rank+, baseado num modelo de inteligência artificial mais antigo do que o DiagAI Score, continua disponível e é aplicado por defeito a todas as análises efectuadas antes de 2 de outubro de 2024.

Baseia-se em 4 componentes principais:

- Uma nova implementação da classificação ACMG, baseada na aprendizagem automática treinada em ClinVar (Score UP²)
- O fenotipo do paciente (se os termos HPO estiverem associados à amostra)
- A concordância entre o genótipo observado e o modo de transmissão da patologia descrito nas bases de dados GenCC, OMIM e "PanelApp Australia".
- Qualidade variante

VariAction Rank. O algoritmo VariAction Rank para a priorização de variantes somáticas é proposto nas análises SomaVar a partir da versão v1.8. Este sistema utiliza a anotação JaxCKB para calcular as pontuações de priorização de variantes com base na sua capacidade de ação clínica. Estas pontuações de priorização têm em conta o impacto da variante, a sua patogenicidade utilizando a pontuação ComPerMed, o AMP Tier, a especificidade da variante com o perfil molecular, bem como a correspondência com a indicação, se a doença foi introduzida antes do lançamento da análise.

ACMG. As variantes podem ser ordenadas de acordo com a sua classificação ACMG selecionando esta opção. Por defeito, as variantes são ordenadas por ordem decrescente da classe «Pathogenic 5» para a classe «Benign 1».

Locus. As variantes podem ser ordenadas de acordo com a sua posição genómica selecionando esta opção.

Allele frequency. As variantes podem ser ordenadas por frequência alélica selecionando esta opção. Por defeito, as variantes são ordenadas por ordem descendente. É possível ordenar por

ordem ascendente clicando no ícone de ordenação . 

GnomAD Allele Frequency, GnomAD Allele Count. As variantes podem ser classificadas pela sua frequência alélica («Allele Frequency»: AF) ou contagem alélica («Allele Count»: AC) no GnomAD, em ordem ascendente ou descendente (se os dados GnomAD estiverem disponíveis).

UP². As variantes podem ser ordenadas pela pontuação de patogenicidade UP², uma componente do DiagAI Score.

Gene. As variantes podem ser ordenadas por ordem alfabética ascendente ou descendente de acordo com o símbolo do gene.

Para variantes que afetam genes sobrepostos, o símbolo do gene utilizado para a ordenação é o associado à variante com maior impacto, com base nas regras de priorização de transcritos descritas acima. Ao expandir a lista de transcritos, o utilizador pode ver o segundo gene sobreposto, que pode aparecer não ordenado. A ordenação também pode parecer não ser aplicada quando múltiplos filtros são aplicados à tabela de variantes.

Configuração das colunas da tabela de variantes

As colunas apresentadas no quadro de variantes podem ser configuradas a partir do menu

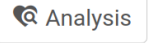
"Organise your columns", acessível através do botão  no canto superior direito do quadro de variantes. A apresentação das colunas pode ser modificada pelo utilizador clicando nos ícones  para ativar a(s) coluna(s) pretendida(s) na secção esquerda. A secção direita apresenta as colunas activas na análise e permite aos utilizadores criar e guardar perfis de colunas a partir de um menu pendente, clicando no ícone "Select profile".

As colunas activas podem ser reordenadas clicando e arrastando-as para a localização pretendida. Por defeito, as colunas são organizadas de acordo com os perfis "Germline" e "Somatic".

Quando os utilizadores passam o rato sobre o nome de um perfil de coluna no menu pendente de predefinições, uma dica de ferramenta mostra qual o workset que criou esse perfil. Isto avisa os utilizadores se o seu perfil pode ter colunas incompatíveis com o workset atual.

Para criar um novo perfil de coluna, basta introduzir um nome no campo "Create a new preset" e clicar em "Save".

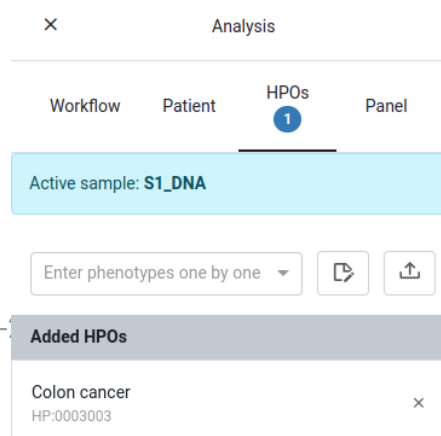
Editar os metadados da amostra a partir do painel de análise

O utilizador pode editar os metadados da amostra diretamente a partir da análise e introduzir o sexo do paciente ou HPOs (apenas para análises de linha germinativa) a partir do "Analysis Drawer", acessível clicando no botão  no canto superior direito da página (ver secção **Painel lateral para gestão de análises**).

A partir da vista Patient, o utilizador acede à edição do género (ver a secção **Patient modal view**).

Nas análises de projetos Germline, a partir da vista de HPOs, o utilizador tem acesso a várias funções de edição de termos de HPOs:

- Ver os termos já introduzidos antes de lançar a análise
- Adição manual de códigos HPO ou fenótipos através da barra de pesquisa "Enter phenotypes one by one".



The screenshot shows the 'Analysis' drawer with tabs for Workflow, Patient, HPOs (selected), and Panel. Below the tabs, it says 'Active sample: S1_DNA'. There is a search bar labeled 'Enter phenotypes one by one' with a dropdown arrow and two icons (a magnifying glass and an upload icon). Below the search bar, there is a list titled 'Added HPOs' with one entry: 'Colon cancer' with the code 'HP:0003003' and a close button (X).

- Importar um ficheiro de texto com códigos HPO
- Geração de termos HPO com base na descrição clínica do doente, acedendo ao DiagAI" através do ícone  . Esta funcionalidade pede ao utilizador que descreva o perfil clínico do doente em Italian (descrição limitada a 4000 caracteres). De seguida, após clicar em "Extract HPOs with DiagAI", é gerada uma lista de HPOs destacados a cores. Finalmente, o utilizador pode associar todos os termos HPO propostos ao paciente clicando em "Add all", ou seleccioná-los individualmente passando o rato por cima dos termos escolhidos e clicando em "Add".

A adição e edição de HPOs a partir desta vista dá ao utilizador a opção de reprocessar a pontuação DiagAI e a classificação da variante (Short list ou SmartPick) de acordo com os novos HPOs introduzidos através da funcionalidade "Re-process analysis".

Variantes de filtragem

É possível refinar as pesquisas no separador *Variants* utilizando as categorias de filtros detalhadas abaixo. O painel de filtros pode ser acedido através do ícone *Filters*  localizado no canto superior esquerdo da tabela de variantes ou premindo a tecla F no teclado. Os filtros estão disponíveis no separador "Basic". Na parte superior do painel de filtros, a secção "Active Filters" mostra o número de filtros atualmente aplicados.

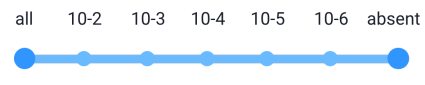
Filtros Genes & Variants

Nome do filtro e visão geral	Descrição
Genes ⓘ Search gene symbols	Restringe a lista de variantes a um ou mais genes.
Variants ⓘ Search by HGVS	Restringe a lista de variantes a uma ou mais variantes de acordo com a nomenclatura HGVS ou HGVS.

Secção Samples & Quality

Nome do filtro e visão geral	Descrição
Coverage ⓘ 1 1 836	Define um limiar mínimo para a profundidade de cobertura (COV). O valor máximo apresentado por este filtro corresponde ao valor de cobertura mais elevado identificado entre as variantes.
Allele Frequency ⓘ 0% 0 100 100%	Filtra as variantes de acordo com a frequência do alelo mutante na amostra analisada.
Quality ⓘ Low Medium High Low Medium High	Qualidade média (pontuação phred) de cada base portadora do alelo mutado, indicando a confiança na deteção da variante: - <i>Medium</i> ($\geq \text{QUAL}=20$): 1 hipótese em 100 de chamada incorrecta, - <i>High</i> ($\geq \text{QUAL}=30$): 1 hipótese em 1000 de chamada incorrecta. Este filtro aplica-se a todas as amostras.
Quality ⓘ All Low Medium High All Low $\geq Q8$ Medium $\geq Q20$ High $\geq Q30$	Para análises GermVar Tertiary e GermVar Tertiary Family em dados da Oxford Nanopore Technologies, a qualidade baseada na pontuação Phred indica confiança na deteção de variantes. $\geq \text{QUAL}=8$ (Low): 1 hipótese em 8 de chamada incorrecta,

Secção Frequencies

Nome do filtro e visão geral	Descrição
▼ Sample frequency ⓘ 	Elimina as variantes com base na sua recorrência no projeto em estudo, da mais frequente (all) para a menos frequente (rare).
▼ Relative pop frequency ⓘ 	Seleciona variantes de acordo com a frequência do seu alelo mutante na população combinada dos projetos ExAc e gnomAD.
▼ Absolute pop frequency ⓘ 	Seleciona variantes com base no número de alelos na população combinada ExAc e GnomAD.
▼ Absolute homozygous pop frequency ⓘ 	Seleciona variantes com base nas suas frequências de homozigotia na população combinada ExAc e GnomAD.

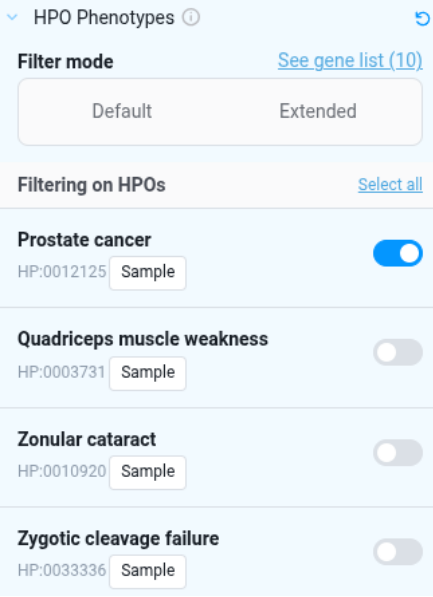
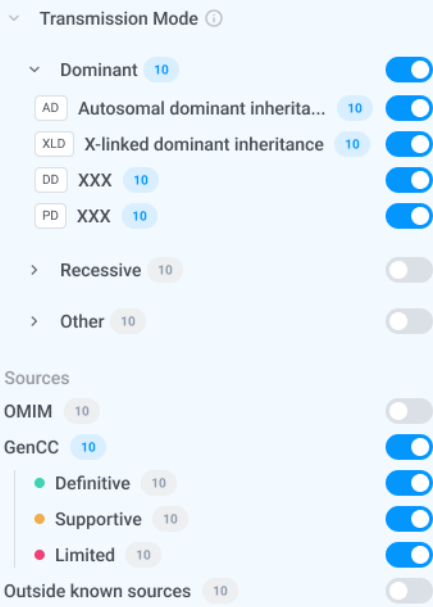
Secção Genetics

Nome do filtro e visão geral	Descrição
Impacto ⓘ Low Moderate High ☒ Keep <i>Modifier</i>	O impacto da mutação na proteína. <i>Modificador:</i> afecta os genes não codificantes
Effects ⓘ > Loss of Function 10 Transcript ablation 10 Splice acceptor 10 Splice donor 10 Stop gained 10 Frameshift 10 Stop lost 10 Start lost 10 Transcript amplification 10	Termos SO (Sequence Ontology) que descrevem a consequência mais perturbadora a nível do gene. Este filtro permite seleccionar o impacto da mutação no gene, e as opções estão organizadas em categorias lógicas para melhorar a navegação: <i>Loss of Function</i>, <i>Coding</i>, <i>Splicing region</i>, <i>Silent variant</i>, <i>Intronic</i>, <i>UTR</i>, <i>ncRNA</i>, <i>Upstream/Downstream</i>, <i>Regulatory</i>, <i>Others</i>.
Gene panel ⓘ Select panel	Selecionar e aplicar um ou mais painéis de genes. Este filtro apresenta todos os painel de genes disponíveis para a entidade. Este filtro não está disponível se tiver sido aplicado um painel restrito à análise a partir do separador <i>Panel (Painel)</i> do painel de gestão da análise (ver secção Vista Modale Panel). Uma mensagem informa o utilizador que o filtro 'Gene panel' não pode ser utilizado com um painel restrito.
Panel in-silico ⓘ Dyslipidemia_panel_1_Twist★	Seleciona e aplica um painel in silico". Este filtro apresenta todos os ficheiros BED disponíveis para o grupo. O nome do ficheiro utilizado durante a análise é realçado com uma estrela. Apenas as variantes cuja posição inicial (start) se situa num dos intervalos do ficheiro de base selecionado são retidas por este filtro.

PanelApp ⓘ Epi ENG Blepharophimosis ptosis and epicanthus ENG Early onset or syndromic epilepsy ENG Epidermodysplasia verruciformis	Permite-lhe filtrar as variantes de acordo com uma lista de painéis in silico nos catálogos PanelApp Australia e PanelApp England". Um emblema identifica a base de dados da qual o painel in silico é originário: AUS" (Austrália) ou "ENG" (Inglaterra). Em ambos os catálogos, os superpainéis não estão disponíveis para seleção, bem como os painéis que não incluem pelo menos um gene. Por exemplo, um painel que contenha apenas STRs não será selecionável.
Gene occurrences ⓘ Usable only if they are fewer than 2000 selected variants.	Define um limiar para o número de ocorrências de um determinado gene na tabela de variantes.
Chromosomes & MT ⓘ X 25 Y 0 M 0	Mantém as variantes localizadas nos cromossomas sexuais e/ou no cromossoma mitocondrial (M)
RefSeq transcripts ⓘ BARD1_2transcripts.txt	Seleciona uma lista de transcrições RefSeq para filtrar as variantes.

Secção Databases

Nome do filtro e visão geral	Descrição
Variant Databases ⓘ dbsnp 31 cosmic 31 civic 30 lovd 10 clinvar 3	Seleciona variantes anotadas pelas seguintes bases de dados: dbsnp, cosmic, lovd, clinvar e civic.

	Permite que as variantes sejam filtradas com base em termos HPO previamente introduzidos, quer antes da análise da amostra (ver Sample metadata), quer após a análise a partir do separador HPO do painel de análise (ver vista modal HPO). As correspondências são identificadas utilizando a nossa abordagem PhenoGenius, que se baseia numa matriz HPO-gene para seleccionar genes. Estão disponíveis 2 opções no modo Filter : - Filtragem predefinida que equilibra a precisão e o número de genes devolvidos. - Filtragem alargada maximizando a precisão com mais genes. A opção "See gene list" fornece uma lista de genes associados aos HPOs seleccionados neste filtro. Esta funcionalidade só está disponível se os termos HPO estiverem seleccionados. Abre uma janela pop-up que mostra a ID, o símbolo, a classificação Phenogenius e a pontuação Phenogenius de cada gene. Pode descarregar a lista de genes utilizando o ícone Download ou copiá-la utilizando o botão copiar/colar.
	Permite aos utilizadores filtrar variantes utilizando ligações de herança de doenças relacionadas com a genética médica humana. As variantes podem ser filtradas de acordo com os modos de transmissão (Dominant (dominante), Recessive (recessivo), Other (outro)) e as fontes de transmissão (OMIM, GenCC ou outras bases de dados).

Transmission mode ⓘ [Mi] Mitochondrial inheritance 57 [AD] Autosomal dominant inheritance 166 [AR] Autosomal recessive inheritance 168 [SMO] Somatic mosaicism 2 [Mu] Multifactorial inheritance 7 [SMu] Somatic mutation 42 Show All	Especifica o modo de transmissão para o(s) fenotipo(s) associado(s) à variante, consultando as bases de dados OMIM e GenCC. Os utilizadores podem clicar na dica de ferramenta deste filtro para alternar entre "Keep unknown" (Manter desconhecido), que determina se devem ser incluídas variantes sem modos de transmissão associados (N/A ou unknown).
Transmission pattern ⓘ Compound Het - Suspected 10 ☒ Compound Het - Verified 10 ☒ Recessive Hom 10 ☒ De Novo 10 ☒	Permite a seleção de variantes de acordo com os modos de transmissão heterozigótico composto (suspeito ou verificado), homozigótico recessivo ou de novo, com base no DiagAI. Tenha em atenção que os filtros "Compound Het - Verified", "Recessive Hom" e "De Novo" só estão disponíveis em análises de famílias.

Secção Classification

Nome do filtro e visão geral	Descrição
▼ VKB ⓘ New Unknown variant 5 Pathogenic 4 Likely Pathogenic 3 Uncertain Significance 2 Likely Benign 1 Benign Possible artifact ☐ Analysis only ⓘ	Permite-lhe selecionar variantes desconhecidas não avaliadas (New - Unknown variant) ou validadas no VKB em todas as análises da entidade. Se a caixa 'Analysis only' estiver selecionada, as variantes serão pesquisadas apenas na análise que está a ser anotada atualmente.
▼ ACMG ⓘ 5 Pathogenic 5 4 Likely Pathogenic 1 3 Unknown significance (VUS) 56 2 Likely Benign 49 1 Benign 6	Pathogenicity classification' para variantes germinativas, calculada numa escala de 1 (benigna) a 5 (patogénica).
▼ ComPerMed ⓘ 5 Pathogenic 7 4 Likely Pathogenic 6 3 VUS 102 2 Likely Benign 1 1 Benign 7	Seleciona as variantes somáticas de acordo com a sua classe de patogenicidade ComPerMed numa escala de 1 (benigna) a 5 (patogénica).
▼ Evidence Type (CKB) ⓘ Actionable 1346 Diagnostic 0 Prognostic 0	Seleciona variantes de acordo com 3 tipos de provas (CKB): A (Actionability), Diagnóstico, Prognóstico.

AMP/ASCO/CAP Tier ⓘ I II 1346 0	Selecione variantes de acordo com o nível AMP que lhes foi atribuído (níveis I, II e N/A) apenas na análise do contexto somático.
---	---

Secção Scores

Nome do filtro e visão geral	Descrição
CI-SpliceAI ⓘ all > 0.2 > 0.5 > 0.8 Keep unknown	Selecione pontuações delta. A caixa "Keep unknown" mantém as variantes com um valor de "N/A", mas exclui as variantes sem impacto de splicing (variante "No splicing impact").
Conservation ⓘ Low Conserved Highly Conserved	Grau de conservação do sítio, calculado a partir de múltiplas pontuações de conservação.
Metasvm ⓘ Tolerated Damaging Highly Damaging	Tem em conta a pontuação de 10 aplicações de previsão de variantes e a frequência máxima observada na população 1000genomes.
Universal Pathogenicity Predictor (UP²) ⓘ Benign VUS Pathogenic -1 0 1 Min -1 Max 1	Seleciona as variantes de acordo com o preditor universal de patogenicidade (UP²).
LOEUF ⓘ Intolerant Tolerant 0 1 2 Min 0 Max 2 Keep unknown	Seleciona as variantes de acordo com a pontuação LOEUF.

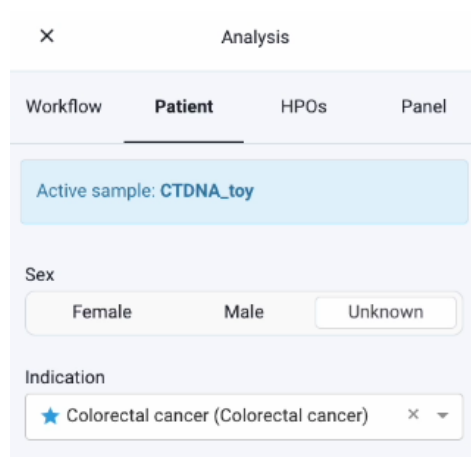
▼ CADD ⓘ Benign Pathogenic 1 99 Min 1 ^ ▼ Max 99 ^ ▼ ☒ Keep unknown	Seleciona as variantes de acordo com a pontuação CADD Phred.
▼ REVEL ⓘ Benign Pathogenic 0 0.5 1 Min 0 ^ ▼ Max 1 ^ ▼ ☒ Keep unknown	Seleciona as variantes de acordo com a pontuação REVEL.

Selecionar uma doença

No SomaVar, SomaVar LF e SomaCGP, é possível selecionar uma indicação de patologia predefinida da base de dados JaxCKB a partir de um menu pendente disponível no separador *Patient* da barra lateral de gestão de análises (ver secção **Patient Modal View**).

O sistema de ontologia baseia-se na *Disease Ontology*²⁴ e inclui termos específicos do JaxCKB. A informação de sensibilidade, na subcoluna "A" da coluna "Clinical Evidences" (ver secção **Variant Page**), é realçada a verde apenas quando a indicação para essa variante corresponde à indicação selecionada. Se forem encontradas asserções de resistência para uma variante, a célula correspondente na coluna 'A' será destacada a vermelho, independentemente da indicação. Caso contrário, ficará a cinzento.

Se a doença tiver sido selecionada durante o carregamento da amostra e antes da análise (ver **Sample metadata**), torna-se a indicação predefinida definida para a amostra e no separador *Patient* (*Paciente*) do painel lateral de gestão da análise. É marcada com um ícone de estrela e aparece em primeiro lugar na lista. A alteração da indicação no painel lateral de gestão da análise não actualiza a doença pré-selecionada para a amostra (Metadados da doença).



The screenshot shows a web interface for an 'Analysis' modal. At the top, there's a close button (X) and the title 'Analysis'. Below this is a tabbed interface with four tabs: 'Workflow', 'Patient', 'HPOs', and 'Panel'. The 'Patient' tab is currently selected. Inside the 'Patient' tab, there's a light blue box stating 'Active sample: CTDNA_toy'. Below that, there's a 'Sex' section with three buttons: 'Female', 'Male', and 'Unknown'. The 'Female' button is highlighted. Underneath is an 'Indication' section with a dropdown menu. The selected item is '★ Colorectal cancer (Colorectal cancer)', which is marked with a blue star icon. There are 'x' and 'v' icons next to the dropdown text.

²⁴ <https://disease-ontology.org/>

Utilizar um perfil de filtros predefinido

Os utilizadores podem seleccionar um perfil de filtro predefinido a partir de um menu pendente, clicando no ícone "Select profile". A plataforma SeqOne oferece vários perfis de filtro predefinidos por defeito: *GermVar*, *My Selection*, *Research*, *DiagAI*.

Editar e guardar perfis de filtros

Os utilizadores podem guardar a posição atual dos cursores para cada filtro como um perfil de filtro, dando um nome ao perfil e clicando em "Save as new". O perfil de filtro estará então disponível através do mesmo menu pendente na categoria "Private".

Ao guardar um perfil de filtro, abre-se um formulário de criação e o utilizador pode seleccionar o botão com o nome da entidade para a tornar acessível a outros utilizadores da entidade. Se o utilizador seleccionar o botão "Private", o perfil só estará acessível ao utilizador que o criou.

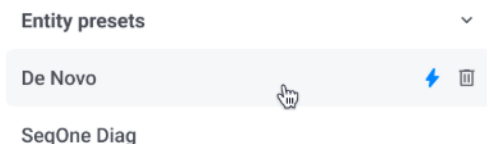
Só o utilizador que criou inicialmente o perfil de filtro poderá modificá-lo clicando em "Update"; se o perfil for modificado por outro utilizador da entidade, ser-lhe-á pedido que o guarde com um nome diferente. Se o utilizador não tiver autorização para modificar um perfil de filtro público, ser-lhe-á pedido que contacte o administrador da sua entidade. Os perfis de filtro SeqOne não podem ser modificados e guardados sob outro perfil de filtro.

É possível repor um ou mais filtros. Basta clicar no ícone  à direita do filtro ou no canto superior direito da caixa "Filters".

Perfis de filtros favoritos

Os utilizadores podem definir perfis de filtro como favoritos, clicando no ícone  à direita do perfil, para que apareçam como um atalho diretamente na barra de acesso rápido acima da tabela de variantes.

Para remover um perfil de filtro dos favoritos, basta clicar no mesmo ícone . Por defeito, o perfil *My Selection* é definido como favorito.

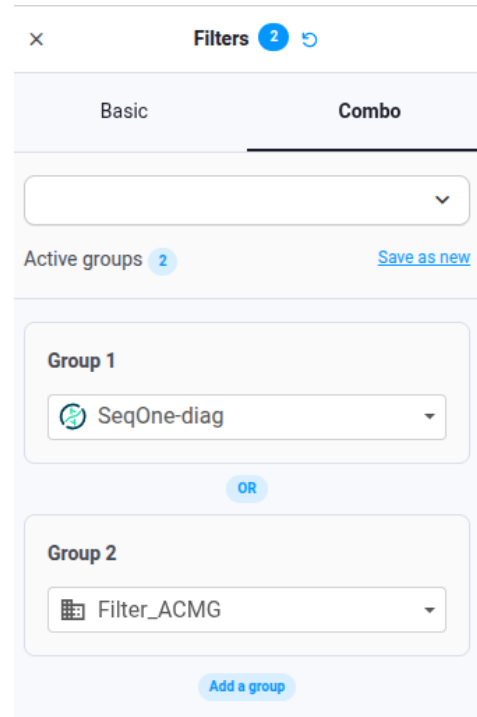


Combinação de perfis de filtro

Os perfis de filtro podem ser combinados sob a condição **OR**, clicando no separador Combo no painel de filtros e, em seguida, em "Create a combo". Por defeito, o formulário de criação mostra 2 grupos de perfis de filtro, mas os utilizadores podem adicionar mais clicando em "Add a group".

Uma vez seleccionados os perfis, os utilizadores são convidados a guardar o seu perfil combinado clicando em "Save as new". Tal como acontece com os perfis de filtro, o utilizador pode tornar o seu combo acessível a todos os utilizadores da entidade no formulário de criação. Apenas o utilizador que criou inicialmente o perfil de combinação poderá editá-lo clicando em "Update".

Se o perfil for modificado por outro utilizador na entidade, ser-lhe-á pedido que o guarde com um nome diferente. Se o utilizador não tiver permissão para modificar um perfil combo público, ser-lhe-á pedido que contacte o administrador da sua entidade.



Perfis de filtros SeqOne

My Selection. Este perfil de filtro permite que as variantes avaliadas como 3 (*VUS*), 4 (*Likely Pathogenic*) e 5 (*Pathogenic*) sejam apresentadas diretamente no VKB na análise que está a ser interpretada.

DiagAI. Este perfil de filtro, disponível nas análises germline (GermlineVar, GermlineFamily, GermVar Tertiary, GermVar Tertiary Family), apresenta uma lista reduzida de variantes com uma elevada probabilidade de serem a causa da doença do paciente. Este perfil explora o algoritmo de priorização DiagAI e a ocorrência da variante no projeto (filtro Sample Frequency).

Pode ser aplicado quando os termos HPO estão ou não associados à amostra, mas a sua sensibilidade é aumentada na presença de termos HPO. No caso de uma análise de família, o perfil do filtro adapta-se automaticamente ao tamanho da família analisada, pois tem em conta que a ocorrência de variantes no projeto pode ser maior. Os utilizadores podem também modificar o filtro Sample Frequency no projeto em qualquer altura.

Introduzir HPOs: para apresentar as variantes SmartPick com maior probabilidade de causar doença.

Family profiles. Nas análises GermlineFamily e GermVar Tertiary Family, os perfis *De Novo (New)*, *Recessive Hom (New)* et *Compound Het (New)* estarão disponíveis apenas para análises de trios que incluam um caso índice afetado, o pai não afetado e a mãe não afetada. Caso contrário, estes perfis serão desactivados.

Não é recomendável modificar os perfis *De Novo (New)*, *Recessive Hom (New)* et *Compound Het (New)* ao criar e guardar um perfil pessoal. Recomenda-se que se comece do zero quando se cria uma variante destes perfis.

Carrier screening. Este perfil de filtro só está disponível no GermlineFamily para análises duo (pais sem amostra índice) e identifica genes recessivos com uma variante heterozigótica potencialmente patogénica em ambos os pais "não afectados" para testes pré-concepcionais.

Somatic FFPE. Este perfil de filtro apresenta uma lista reduzida de variantes somáticas em análises somáticas (SomaVar, SomaCGP) efectuadas em ADN FFPE.

Somatic LF. Este perfil de filtro apresenta uma lista reduzida de variantes somáticas nas análises SomaVar LF.

Função "Mark as viewed"

"Mark as viewed" (Marcar como visualizado) é uma característica que permite ao utilizador ocultar seletivamente certas variantes. Para isso, basta seleccionar a(s) variante(s) desejada(s)

e clicar no ícone  Mark as viewed exibido na parte inferior da tabela de variantes. A seleção de variantes aparece então a cinzento. Para remover a seleção, o utilizador pode clicar no ícone

 Reset all viewed na parte superior da tabela de variantes para repor a vista, ou seleccionar a(s) variante(s) e clicar em  Mark as not viewed.

O utilizador pode ocultar o pop-up clicando na cruz. Isto anulará a seleção de todas as variantes, mas estas permanecerão a cinzento.

Selecionar variantes

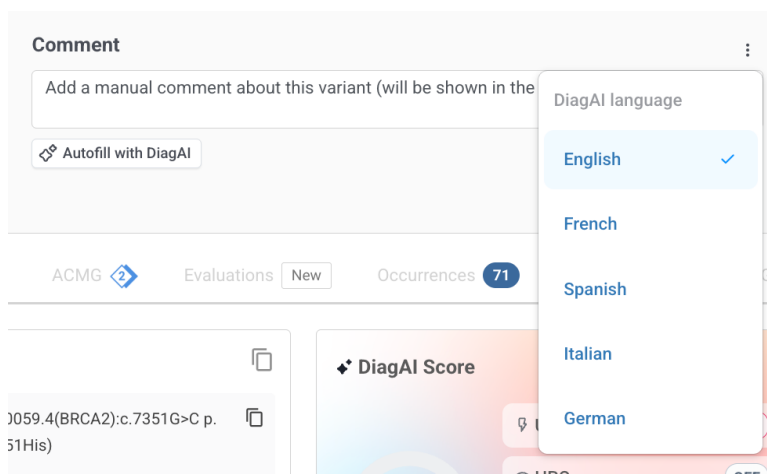
As variantes podem ser adicionadas ao painel "Selection" juntamente com quaisquer variantes avaliadas (ver secção abaixo). Para tal, selecione as variantes marcando as caixas do lado esquerdo da tabela de variantes e, em seguida, clique no ícone "Add to selection" na parte inferior da tabela.

Avaliação de variantes

As variantes podem ser rapidamente avaliadas a partir da tabela de variantes, clicando na coluna "VKB" e especificando a sua classe de patogenicidade e um comentário opcional. Estas avaliações e comentários estão acessíveis a todos os utilizadores da plataforma dentro da mesma entidade através do ícone "Selection" na análise (ver secção **Painel lateral para gerir avaliações**). Uma variante também pode ser avaliada a partir da sua página dedicada.

Os comentários podem ser introduzidos manualmente ou gerados automaticamente. A interface inclui um campo vazio "Classification" (Classificação) para os comentários e um botão "Autofill with DiagAI" (Preenchimento automático com DiagAI) para a geração automática. É apresentado um indicador durante a geração de comentários. Para os comentários existentes, a opção "Improve with DiagAI" permite-lhes ser melhorados. Cada comentário gerado pode ser aprovado, modificado ou anulado. Uma vez validados, os comentários são guardados com a sua classe de patogenicidade, data e conteúdo. Podem ser modificados ou eliminados após a validação, mantendo o seu historial.

Para seleccionar o idioma do DiagAI, clique nos três pontos no canto superior direito da caixa de comentários e escolha o idioma de sua preferência no menu suspenso.

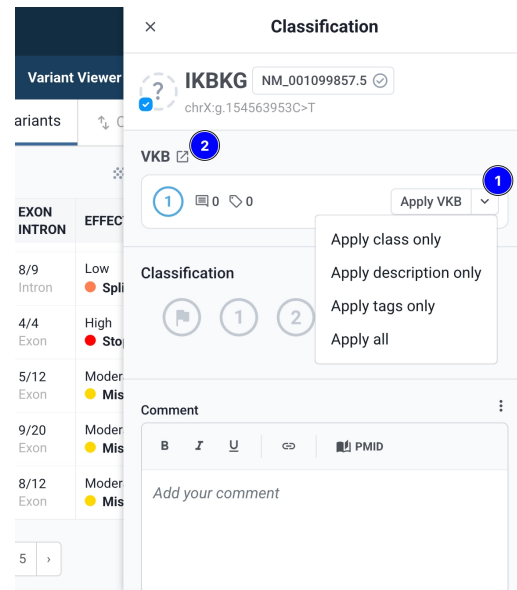


As variantes também podem ser classificadas clicando na coluna de classificação, cujo cabeçalho tem uma bandeira. Para classificar uma variante, clique no ícone +. Abre-se um painel lateral de classificação, permitindo ao utilizador adicionar a classe de patogenicidade, um comentário e uma etiqueta.

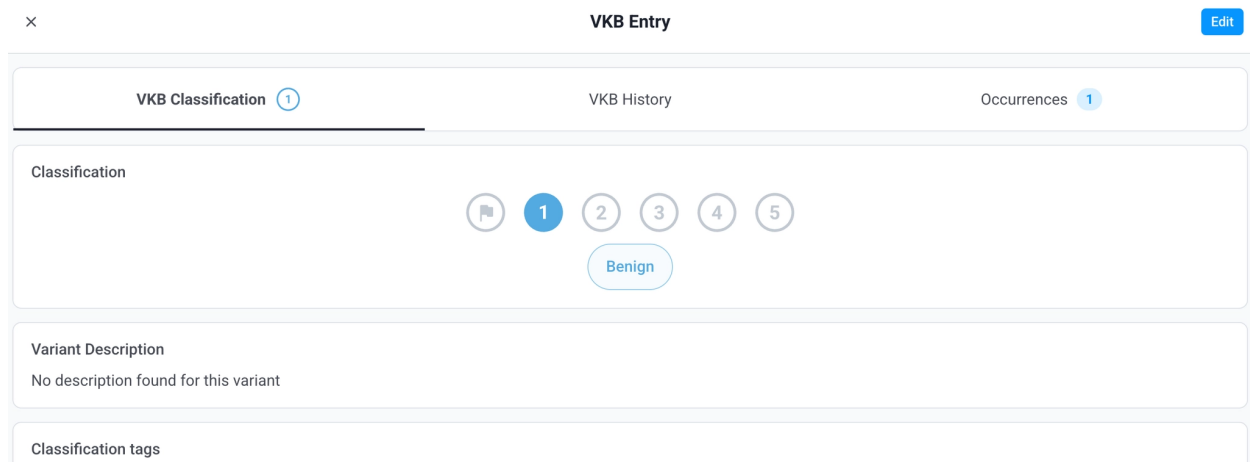
Para adicionar esta classificação à base de dados VKB, ative o botão "Add to VKB". Se já existir uma classificação na base de dados VKB, clique em "Apply VKB" para aplicá-la à entrada atual. Estão disponíveis opções adicionais clicando na seta junto a "Apply VKB" (ver passo 1 na figura abaixo).

Valide a classificação clicando em "Save" no canto superior direito do painel lateral.

Para editar uma classificação diretamente na base de dados VKB, clique no link de redirecionamento VKB (ver passo 2 na figura abaixo) para abrir a janela "VKB Entry".



Na janela "VKB Entry", clique em "Edit" no canto superior direito para modificar a avaliação e, em seguida, clique em "Save" para aplicar as alterações. O histórico VKB e as ocorrências da variante estão acessíveis navegando através dos separadores dedicados.



VKB Search. Está disponível um painel lateral clicando no botão 'VKB' na barra de pesquisa do cabeçalho do SeqOne. Isto permite-lhe procurar variantes anotadas na base de dados VKB por nome do gene, transcript:, anotação HGVS ou identificador SeqOne

(Genome_variant_chromosome_location_REF_ALT). A pesquisa pode ser filtrada por genoma de referência (GRCh37 ou GRCh38) e/ou classe de patogenicidade.

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: Janela VKB Search.

Quando se clica numa variante avaliada, abre-se uma página mais detalhada com informações sobre a variante, tais como a sua posição genômica, as suas anotações HGVS no ADN complementar (cDNA) e na proteína, as classes de patogenicidade, as análises e os respectivos comentários associados em que foi avaliada e, em seguida, uma ligação para as análises.

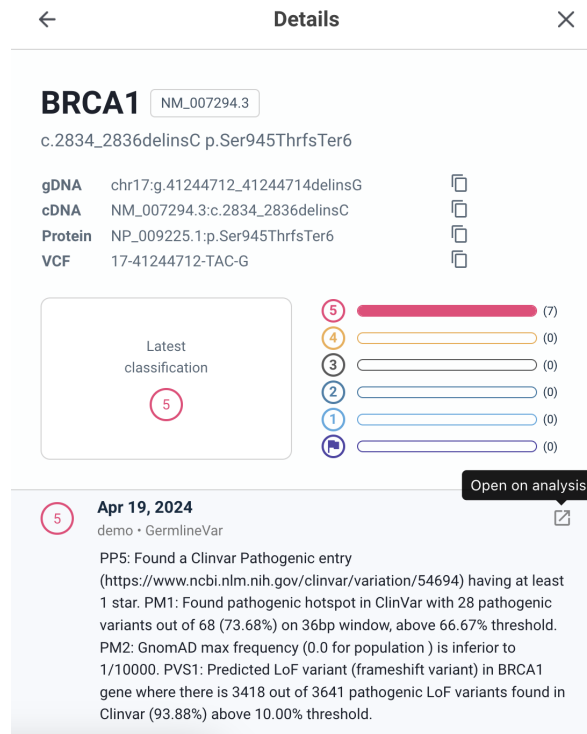


Figura 16 : Janela de detalhes da variante no VKB.

Página de variantes

Clicar numa variante da tabela de variantes abre um painel lateral que contém mais informações, com vários separadores. Os utilizadores podem abrir a página da variante noutra separador clicando em "Open in new window" no canto superior direito. O painel lateral pode ser fechado com a tecla "Esc".

Os utilizadores podem avaliar e comentar uma variante nesta janela. Tal como as avaliações e comentários efectuados a partir da tabela principal de variantes, estas avaliações e comentários podem ser vistos por todos os membros do grupo que utilizam a plataforma.

A partir desta página, os utilizadores podem também aceder a informações sobre variantes, incluindo genótipo (GT), frequência alélica (VAF), profundidade de cobertura (COV) e 'Scores' (QA).

Os separadores da página de variantes são descritos abaixo.

Variant. O separador está dividido em vários blocos, descritos abaixo.

- **Nomenclatura:** são enumeradas as diferentes nomenclaturas (g., c., p., entrada VCF), com a possibilidade de copiar as informações de todo o bloco ou de cada linha para a área de transferência utilizando o ícone .
- **DiagAI Score:** utiliza o bloco de resumo do cálculo do DiagAI Score disponível no separador DiagAI Score na página da variante (ver parágrafo **DiagAI Score**).
- **ClinVar:** informações retiradas da coluna ClinVar Germline da tabela de variantes, indicando a classe de patogenicidade atribuída à variante, o seu status clínico e o número de entradas na ClinVar (ver **Visualização e Interpretação** na secção **Variantes Pequenas**). O ícone  abre a página ClinVar num novo separador.
- **Anotações:** impacto do transcript, exão em causa, distância do sítio de splicing mais próximo e tipo de mutação.
- **Transmission:** ligações às bases de dados OMIM e GenCC se uma ou mais doenças estiverem associadas ao gene.
- **Compound Heterozygous Candidates**, este bloco está dividido em 2 partes:
 - A secção "*Short Variants*" agrupa todos os SNVs candidatos localizados no mesmo gene (com base na ID do gene) e classifica-os de acordo com a sua pontuação DiagAI. A variante seleccionada na página Variantes é apresentada em primeiro lugar e está a cinzento na lista.
 - A secção *CNVs* indica se foram encontrados CNVs, ordenados por classe ACMG. Esta secção só está disponível se tiver sido efectuada uma análise CNV; caso contrário, é apresentada uma mensagem de aviso com a indicação '*CNVs not available*'.

Ao clicar num dos resultados, abre-se um novo separador com a página da variante (se for uma variante pequena) ou a página da CNV (se for uma variante de cópia). Se não forem encontrados CNVs, é apresentada uma mensagem de aviso com as palavras "no CNVs found".

- **Tools:** ligações a outras bases de dados (LitVar, LOVD, VarSome, dbSNP, Cosmic, GnomAD, CKB, UCSC, Decipher, Ensembl).
- **GnomAD:** se a variante for conhecida nesta base de dados, é apresentado um quadro que resume a frequência da variante em diferentes populações e é fornecido acesso direto ao GnomAD através do ícone .

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

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

Gene database

[OMIM](#)
[Ensembl](#)
[NCBI](#)

Genome database

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

Figura 17: Separador Variant na página de variantes

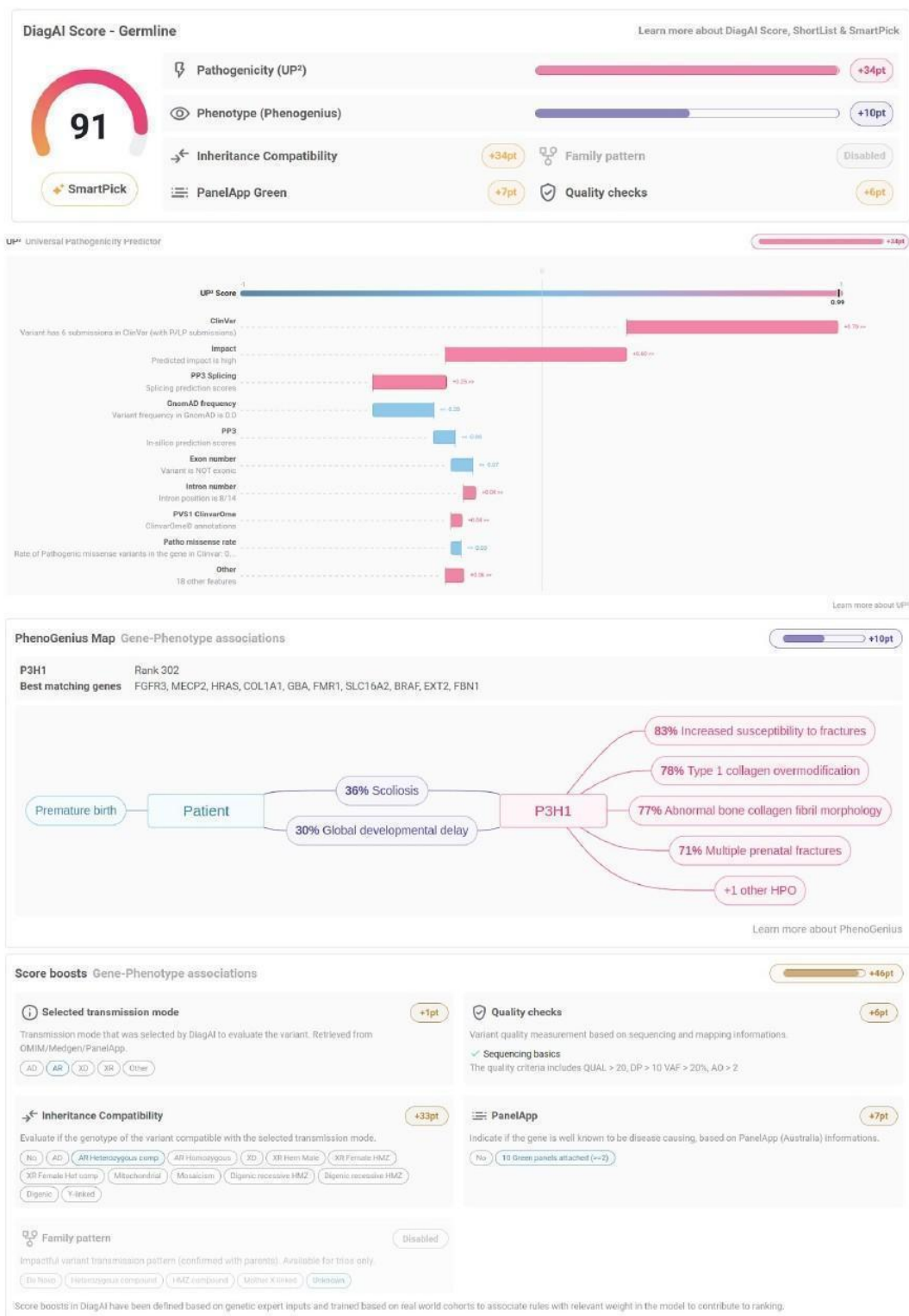


Figura 18: Visualização do separador DiagAI" na página da variante

DiagAI. O separador contém explicações sobre a pontuação DiagAI atribuída à variante. Este separador só está disponível para análises GermlineVar, GermlineFamily, GermVar Tertiary e GermVar Tertiary Family (a partir da v1.8).

O separador DiagAI está dividido em 4 secções principais:

- Um bloco de resumo **DiagAI Score - Germline** que mostra a pontuação, o estado SmartPick e enumera as principais características que contribuem para a pontuação global e o número de pontos associados.
- Um gráfico em cascata que explica o score **UP² (Universal Pathogenicity Predictor)**. É apresentada uma mensagem de aviso  quando o utilizador consulta a página de uma variante cujo transcript difere do utilizado para calcular o UP².
- Um bloco de explicabilidade do **PhenoGenius Map**, colorido se os termos HPOs estiverem associados à amostra. O bloco detalha a classificação do gene, uma lista dos 10 genes mais bem associados e o mapa PhenoGenius, codificado por cores de acordo com a correspondência entre os HPOs do paciente e o gene: "blue" para os HPOs do paciente apenas, roxo para os HPOs do paciente e correspondentes ao gene, vermelho para os HPOs do gene apenas. A secção fica a cinzento se não forem introduzidos termos de HPO para o paciente.
- Uma secção **Score boosts** que integra 5 subcaracterísticas que podem ou não contribuir (Desativado) para a pontuação: Modo de Transmissão, Compatibilidade de Herança, Padrão de Família, PanelApp e Controlo de Qualidade.

ACMG. O separador permite aos utilizadores ver e rever os critérios utilizados durante a classificação automática do ACMG. Podem modificar a seleção verificando os diferentes critérios necessários e, em seguida, utilizar a classificação para avaliar a variante clicando em "Use for Evaluation". Também é possível modificar o nível de evidência (*Supporting*, *Moderate*, *Strong* e *Very Strong*) para atribuir um número diferente de pontos a cada critério, que será tido em conta no cálculo da pontuação final do ACMG. Cada nível de evidência atribuirá: 1 ponto para *Supporting*, 2 pontos para *Moderate*, 4 pontos para *Strong* e 8 pontos para *Very Strong*. Os pontos para cada critério serão então somados para atribuir um estatuto ACMG à variante:

- *Pathogenic*: 10 pontos
- *Likely Pathogenic*: 9 a 6 pontos
- *Uncertain Significance*: 5 a 0 pontos
- *Likely Benign*: -6 a -1 pontos
- *Benigno*: a partir de -7 pontos

Para o critério BA1, o utilizador não pode seleccionar o nível de prova, uma vez que se trata de uma aula autónoma.

TP53 protein coding

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

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

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

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

☒ Use ClinGen point-based system

Predicted:
Likely Pathogenic Use for evaluation

Automatic tags
PVS1 | Strong PM2 | Moderate
PP5 | Moderate

First level tags
No selected tags.

Second level tags
No selected tags.

Tags that do not apply
No selected tags.

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

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

Figura 19: Visualização do separador ACMG na página de variantes

TP53. Este separador só está disponível se a variante estiver anotada no gene TP53. Este separador fornece informações sobre as mutações TP53 e o seu significado clínico de acordo com a base de dados IARC²⁵. O separador TP53 está dividido em vários blocos, descritos abaixo.

Um bloco de anotações (*Annotations*) apresenta as seguintes informações associadas à variante:

- Mutation Rate: taxa de substituição associada à posição da variante
- Hotspot (yes/no): presença da variante num sítio de mutação recorrente (hotspot)
- Structural Motif: motivo estrutural afetado pela mutação
- Domain: domínio funcional da proteína afetada pela mutação

O bloco dados funcionais (*Functional Data*) exhibe classificações e informações funcionais associadas à variante:

- DNE Class: efeito dominante-negativo (DNE) na transactivação da proteína p53 de tipo selvagem
- DNElof Class: classificação funcional para a perda de supressão do crescimento e actividades dominantes negativas

²⁵ <https://tp53.cancer.gov/>.

- Transactivation Class: classificação funcional baseada na atividade transcricional global (TA)
- Structure Function Class: previsões funcionais.

Os blocos de detalhes "Graphics" e "Sample Details" fornecem detalhes das amostras referenciadas na base de dados IARC para a variante. O utilizador pode optar por visualizar os dados associados a amostras somáticas ou germinativas, selecionando "Somatic" ou "Germline".

O bloco 'Graphics' apresenta dois gráficos, com menus pendentes, que mostram a distribuição das amostras em função das suas características.

O bloco 'Sample Details' apresenta informações pormenorizadas sobre a amostra sob a forma de um quadro.

[ComPerMed](#). (apenas análise somática). A classe de patogenicidade de cada variante é calculada de acordo com as recomendações do comité de peritos belga ComPerMed (Comissão de Medicina Personalizada). O utilizador tem acesso a informações detalhadas sobre esta classificação a partir de um separador dedicado na página da variante.

[Evidences](#). O separador fornece uma lista de terapias associadas à variante selecionada a partir da base de dados JaxCKB. O utilizador pode percorrer cada linha para encontrar detalhes sobre as terapias ou indicações.

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
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: Visualização do separador *Evidences* na página da variante

Evaluations. Esse registro fornece acesso ao histórico de avaliações e interpretações da variante ao longo do tempo:

- Na análise atual (secção esquerda do separador)
- Nas outras análises efectuadas no seio do grupo (secção direita do separador)

Occurrences. Este separador identifica as outras amostras que contêm a variante, independentemente de qualquer avaliação. Para cada ocorrência, são introduzidos o nome da amostra, a frequência alélica, a ligação à análise relevante e o nome do projeto.

Publications (Publicações). Esta aba fornece acesso à literatura científica associada à variante e ao gene. Clique na sub-aba "Variant" (Variante) ou "Gene" para visualizar as publicações PubMed que fazem referência à variante ou ao gene. Um link PMID redireciona para o registro PubMed e o link de texto completo abre o artigo completo quando disponível.

Scores. O separador disponível para variantes com um missense, efeitos start lost e stop gained. Com exceção destes casos, a pontuação estará a cinzento. Este separador permite visualizar os resultados das pontuações de previsão de várias ferramentas, tais como *REVEL*, *DANN*, *FATHMM*, *FATHMM-MLK*, *LRT*, *MetaLR*, *MetaSVM*, *MutationTaster 2*, *PROVEAN*, *SIFT* e *PLi*. Também dá acesso a pontuações de conservação como *GERP++*, *PhastCons100way Vertebrate*, *PhastCons20way Mammalian*, *PhyloP100way Vertebrate*, *PhyloP20way Mammalian* e *SiPhHy*.

IGV. Este separador permite ao utilizador visualizar as reads alinhadas com a posição da variante através da versão web do IGV Genome Browser. Em áreas de alta cobertura de leituras, as leituras são subamostradas por padrão, ou seja, nem todas as leituras são exibidas. São aplicados os parâmetros de visualização predefinidos, ou seja, um máximo de 100 reads por janela de 50 nucleótidos. Estas regiões podem ser identificadas pela linha preta por baixo do perfil de cobertura (downsampling). Clicando nela, é apresentado o número de reads removidas

da visualização (# downsampled). Está disponível um menu no ícone  , que oferece três opções:

- *Open in new tab* (Abrir num novo separador): abre a vista da variante IGV num novo separador.
- *Track setting*: permite-lhe adicionar à vista outras amostras do mesmo projeto ou de projetos diferentes.
- *Open local IGV*: redirecciona a aplicação IGV Desktop para as coordenadas da variante que está a ser interpretada. Para visualizar SNVs ou outras variantes, o BAM da amostra deve primeiro ser carregado no aplicativo IGV Desktop. Esta opção só é compatível com as versões do IGV Desktop ≥ 17 .

O IGV suporta vistas divididas, por exemplo, copiando dois conjuntos de coordenadas para a barra de pesquisa de locus, separados por um espaço. Estão disponíveis várias definições clicando nas rodas dentadas à direita do Navegador de Genomas. Por exemplo, para visualizar diretamente reads não alinhadas (soft clips), selecione a opção "Show soft clips".

Gene. O separador contém informações adicionais sobre o gene e a bibliografia associada.

Geração de um relatório clínico

O usuário pode editar um relatório clínico exportável no formato PDF e .docx clicando no botão "Generate report" (Gerar relatório) na aba "Selection" (Seleção) enquanto navega na análise.

Na parte superior da página de edição, o usuário pode selecionar um modelo, definir o status do relatório como novo ou rascunho, selecionar uma amostra (várias amostras podem ser selecionadas para análises familiares) e selecionar o idioma de edição. O usuário também pode alternar entre a visualização "Edition" (Edição) e a visualização "Preview" (Visualização) antes de salvar um rascunho ou prosseguir para a próxima etapa (usando os botões "Save Draft" e "Next Step").

Os modelos de relatório podem ser criados na seção "Reporting", em "Template list" nas configurações da entidade (Entity settings).

Na visualização de edição, o usuário pode editar as diferentes seções do relatório e selecionar várias opções. Na visualização de pré-visualização, o usuário pode revisar o relatório antes de continuar a edição, salvar o rascunho ou validar o relatório.

O usuário pode salvar um rascunho e acessá-lo novamente a partir da aba "Selection", depois "Generate report", como ao criar um novo relatório. Quando um relatório em PDF é gerado, ele é exibido automaticamente na aba "Reporting" (Relatórios) da análise. Somente após a validação o relatório pode ser exportado como .PDF ou .DOCX. O usuário pode excluir o relatório ou criar relatórios adicionais para a mesma análise.

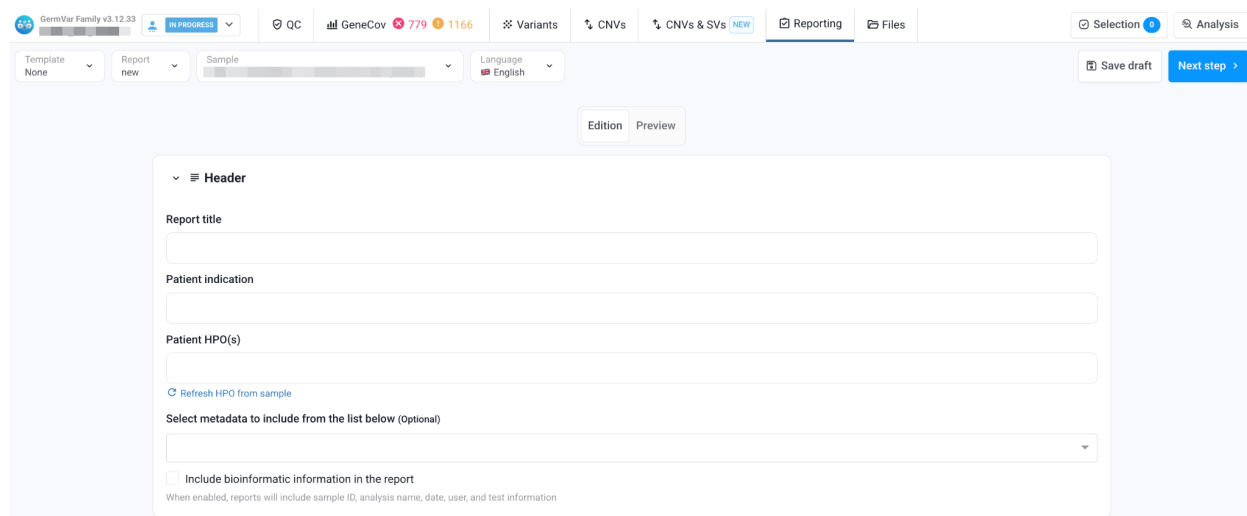
The screenshot shows the 'Reporting' tab of a clinical analysis tool. At the top, there's a navigation bar with tabs like 'Selection', 'Analysis', 'Reporting', and 'Files'. Below this, a header section contains dropdown menus for 'Template' (set to 'None'), 'Report' (set to 'new'), 'Sample', and 'Language'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

Figura 21: Ver o separador do relatório clínico.

13.1.5. CNV

13.1.5.1. CNV em painéis de genes

Metodologia

A detecção de CNVs nos dados do painel de genes, em GermlineVar, GermVar/GermVar Family, SomaVar e SomaHemato, baseia-se na seleção de amostras de controlo da coorte co-analisada, de acordo com a sua correlação.

A correlação é calculada a partir de uma tabela de cobertura, que por sua vez é calculada a partir de todas as regiões do manifesto, mas excluindo os cromossomas sexuais.

Os critérios para que uma amostra seja elegível para a análise de CNV são os seguintes

- Para que uma região seja retida na análise CNV, a cobertura mínima é de 50 reads.
- O valor limiar de correlação para a seleção de amostras de controlo é de 0,90 para o contexto da linha germinal e de 0,75 para o contexto somático.
- O número mínimo de controlos exigido para cada amostra é de seis.

É necessário definir o género das amostras (sample metadata) antes de lançar o CNVCapture ou o SomaCNVCapture para efetuar a identificação dos genes localizados nos cromossomas sexuais. Por último, são necessárias pelo menos oito amostras para efetuar uma análise CNVCapture ou SomaCNVCapture no SeqOne.

Os critérios aplicados ao manifesto do utilizador para que um gene do painel seja elegível para a análise de CNV são os seguintes

- As regiões manifestas são anotadas a partir do transcrito canónico determinado de acordo com o recurso NCBI MANE Select. Se não estiver disponível um MANE Select transcript, a anotação baseia-se no recurso NCBI RefSeq.
- apenas os exões para os quais existe $\geq 80\%$ de sobreposição entre o intervalo manifesto e o exão são analisados. Se nenhum exão do gene se sobrepuser $\geq 80\%$ com o intervalo manifesto, o gene não será incluído na análise da CNV.
- apenas os intrões para os quais existe $\geq 60\%$ de sobreposição entre o intervalo manifesto e o intrão são analisados.

A análise CNVCapture (versão ≥ 2.4) inclui regiões localizadas 20kb a montante e a jusante dos genes cobertos pelo manifesto, se tais regiões estiverem incluídas no desenho do manifesto. Especificamente, cada um dos 20 segmentos de 1kb que constituem a região de 20kb é analisado se pelo menos 20% do segmento se sobrepuser ao manifesto inicial.

Visualização e interpretação

O separador CNVs apresentará um gráfico que representa a lista de genes nos quais foram identificadas variações de cópia. Este separador está disponível nos worksets descritos na secção Metodologia acima.

A terminologia utilizada para descrever o estado dos genes chamados é a seguinte:

- Gain: o gene é afetado por amplificação parcial ou total
- Loss: o gene é portador de uma deleção parcial ou total
- Normal: o gene não é portador de CNV
- Partially not called: o gene é parcialmente chamado
- Not called: o gene é excluído da análise CNV

O estado "partially not called" é atribuído a um gene quando o seu estado de cópia é normal mas pelo menos um exão não foi chamado (com o estado "failed")

Apesar da sua elegibilidade para a análise de CNV, um gene é excluído (Not called) da análise se :

- o gene está num cromossoma sexual e os metadados relativos ao género não foram preenchidos,
- todos os exões falham de acordo com os critérios especificados abaixo: number of controls:", percentage of variation:". O status "Failed" de um gene é definido apenas no CDS. As regiões intrónicas e a montante/jusante são excluídas se os exões CDS forem chamados.

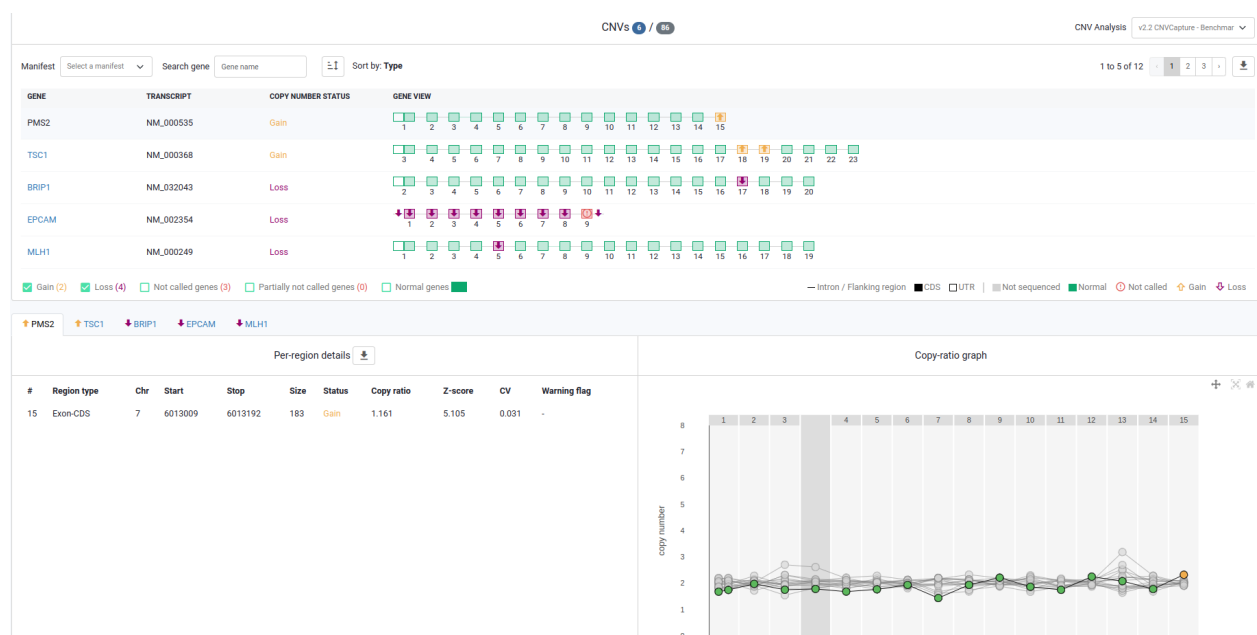


Figura 22: Visualização do separador CNVs numa análise GermlineVar.

Filtros e opções de ordenação

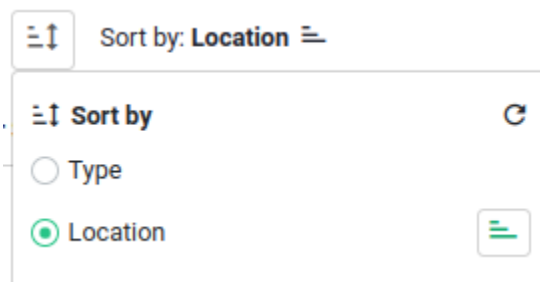
O menu de filtros pode ser utilizado para:

- Restringir a visualização a um subconjunto de regiões/genes utilizando um ficheiro BED que tenha sido previamente carregado na interface com a funcionalidade 'Manifest'
- Procurar um gene específico com a função "Search Gene".

Estão disponíveis duas opções de seleção de CNV:

- **Type** (default): lógica de seleção da corrente (Gain e Loss)
- **Location**: permite ordenar os CNVs" de acordo com os cromossomas e a sua posição inicial, por ordem crescente ou decrescente, clicando

no ícone de ordenação



O utilizador pode também selecionar genes de acordo com o seu estado: Gain, Loss, Normal genes, Partially not called genes ou Not called genes.

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

Detalhes por região

A tabela de detalhes por região mostra os detalhes de cada região para cada Gene:

- #: número da região (exões e intrões)
- **Region type**: tipo de região (a montante, utr-exónica, codificação-exónica, intrónica, a jusante)
- **Chr**: cromossoma
- **Start/Stop**: posição de início e fim da CNV
- **Size**: tamanho da CNV (bp)
- **Status**: status da região (gain = duplicação, lost = deleção, normal = sem CNV, not called = região não foi incluída na análise de CNV)
- **Copy-ratio**: rácio do número de cópias em relação ao normal (n=2).
- **Z-score**: número de desvios-padrão que separam um resultado da média. Valores positivos ou negativos elevados indicam uma forte confiança na identificação da CNV.
- **CV**: coeficiente de variação que mede a dispersão relativa no interior das amostras da coorte.
- **Warning flag**: Ocasionalmente, pode ser apresentada uma mensagem de erro de acordo com os seguintes critérios:
 - **Low zscore**: Valor absoluto do Z-score <4
 - **High variation**: coeficiente de variação (CV) > 0,1
 - **Low coverage**: cobertura da região para esta amostra <100

- **Number of controls (número de controlos):** O número de controlos é insuficiente para poder efetuar a chamada nesta região
- **Percentage of variation:** A variação é demasiado grande entre amostras para que esta região possa efetuar a chamada
- **Read depth:** cobertura insuficiente da região (<50) e rácio de cópias >0,1
- **Missing Gender:** a região está localizada num cromossoma sexual e os metadados relativos ao género não foram introduzidos.

Aviso: uma região rica em Soft clips, como um grande indel ou uma inserção Alu, pode ser interpretada como uma queda na cobertura pelos worksets "CNVCapture" e "SomaCNVCapture" e detetada como uma perda de cópia, se o tamanho da inserção for da mesma ordem que o da região chamada.

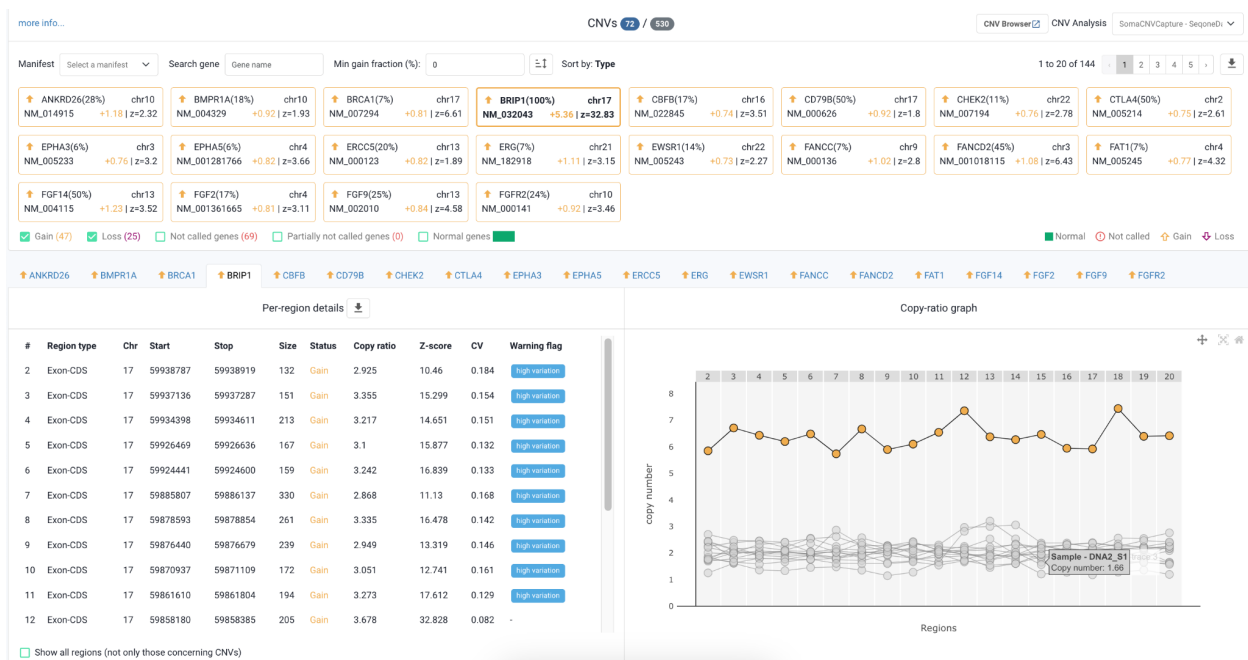
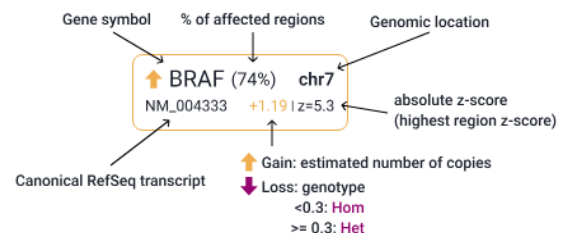


Figura 23: Visualização do separador "CNVs" numa análise SomaVar

No caso das análises **SomaCNVCapture**, embora o método de deteção seja semelhante ao do CNVCapture, a escala utilizada para a representação gráfica das variantes é o gene, e não o exão. A representação é apresentada sob a forma de uma série de caixas. Esta fornece o símbolo do gene, a posição genómica, o transcrito Refseq canónico, o z-score absoluto e, no caso de ganho, o número de cópias ou, no caso de perda, o genótipo.



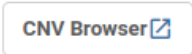
Uma janela pop-up que explica estas caixas está disponível acima dos filtros na secção '[more info...](#)'.

A percentagem apresentada em cada caixa reflecte a fração do gene (*min gain fraction*) afetada pela CNV detetada. A fração pode ser utilizada para filtrar os resultados. Por defeito, o filtro está definido para 0% e aplica-se apenas a ganhos de cópias.

As deleções detetadas por este pipeline não estão sujeitas a este filtro e devem ser interpretadas com cautela. De facto, a capacidade desta ferramenta para detetar variações de uma cópia de um gene depende da natureza e da qualidade da amostra, em particular da fração tumoral.

Aviso: Se tanto uma amplificação como uma deleção forem detetadas no mesmo gene, o gene será considerado como amplificado por defeito e, portanto, sujeito ao filtro "min gain fraction" descrito acima.

CNV Browser

Está disponível um ícone  no separador de visualização de CNV das análises GermlineVar, GermlineFamily, SomaVar ou SomaHemato para abrir um CNV Browser" num novo separador. O gráfico do CNV Browser apresenta a B allele frequency (BAF) de SNPs (Single Nucleotide Polymorphism, polimorfismo de nucleótido único) presentes no manifesto da amostra analisada e presentes num conjunto de dados de matrizes SNP da Illumina e SNPs de outros backbones, como o OneSeq da Agilent, para facilitar a visualização de determinados tipos de eventos, como LoH (Loss of Heterozygosity, perda de heterozigotia).

O BAF é a frequência do alelo não-referência (alelo B') observada para um SNP heterozigótico da linha germinal.

Por defeito, o gráfico representa o BAF das variantes (eixo y) sobre todo o genoma (eixo x), repartido por cromossoma (1-22, X e Y). Um filtro permite ao utilizador visualizar as variantes num determinado cromossoma no canto superior esquerdo.

No contexto das análises de CNV, esta informação BAF pode ser utilizada para mostrar uma alteração no número de cópias específicas de um alelo, perda de heterozigotia ou desequilíbrio alélico.

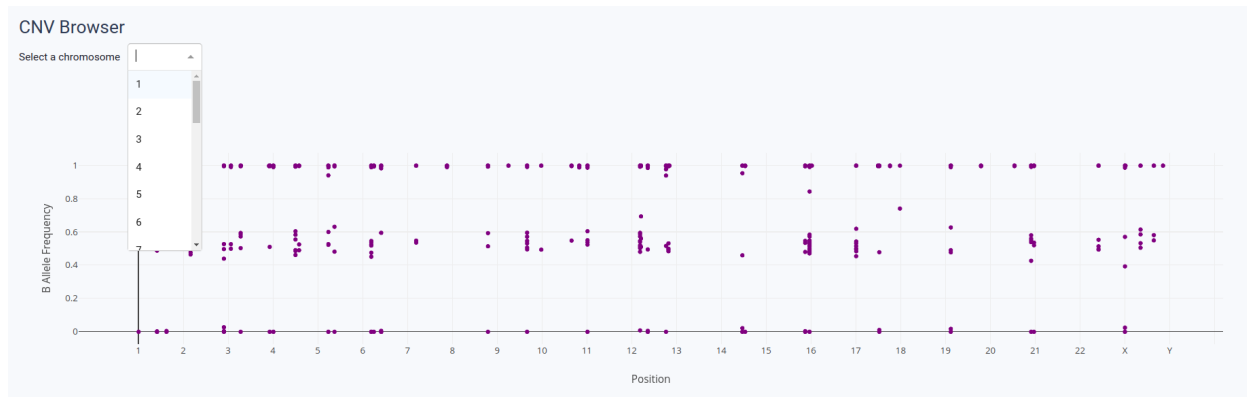


Figura 24: Visualização do separador "CNV Browser" numa análise GermlineVar.

Exportação de dados CNV

O ícone  no canto superior direito pode ser utilizado para exportar os dados da análise CNV sob a forma de um ficheiro .tsv com as seguintes colunas:

- gene (nome do gene),
- RefSeq transcript (refseq_id),
- estado do gene (status),
- cromossoma (chr),
- coordenadas genómicas da CNV: posições inicial e final,
- fração do gene afetado pela CNV (gene_fraction_cnv),
- rácio do número de cópias (copy_ratio) em relação ao normal (n=2),
- número de desvios-padrão que separam um resultado da média (zscore)
- número de exões que se sobrepõem à CNV (exões), no caso de uma análise CNVCapture.

Para cada gene, o ícone  ao lado de "Per-region details" pode ser utilizado para exportar especificamente os dados da análise CNV para o gene selecionado sob a forma de um ficheiro .tsv contendo os campos seguintes:

- gene (nome do gene),
- cromossoma (chr),
- coordenadas genómicas: posições inicial e final,
- número do exão ou do intão (exon_intron_number),
- Status,
- tipo (exão ou intão, tipo),
- mensagens de erro associadas (warning_flags, fail_flags),
- valor crítico de probabilidade ajustado (padj),
- valor crítico de probabilidade (pvalue),
- relação entre o número de cópias e o normal (rácio),
- medida da dispersão relativa dentro das amostras (variabilidade),
- desvio-padrão do resultado em relação à média (zscore)

13.1.5.2. CNV na análise SomaCGP e SomaLBx

Metodologia

O SomaCGP e o SomaLBx utilizam o PureCN²⁶ para detetar CNVs e LOHs (loss of heterozygosity). O PureCN identifica variantes somáticas a partir de dados de detecção de variantes, a fim de avaliar a ploidia e a pureza tumoral da amostra. Esta informação é utilizada para determinar o número de cópias de cada evento.

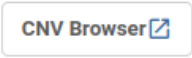
Visualização e interpretação

O separador de visualização de CNV na análise SomaVar Tertiary funciona de forma idêntica à do SomaCGP e do SomaLBx, como descrito abaixo.

Os CNVs são apresentados numa tabela filtrável. Cada CNV apresenta as seguintes informações:

- **Gene:** nome do gene
- **Location:** cytoband
- **Coordenadas do evento / Event size:** coordenadas e tamanho do evento
- **Transcript:** transcrição de referência
- **Exons:** exões afectados
- **Type:** tipo de evento (LoH, Loss ou Gain)
- **Copy number:** número absoluto de cópias
- **Function:** função do gene, se aplicável, Oncogene ou Tumor Suppressor Gene (TSG)

O utilizador pode procurar CNVs que afectem um determinado gene através de um campo de pesquisa e escolher se quer ou não mostrar genes presentes em regiões LOH, marcando a caixa "Show genes in LOH regions". A tabela de resultados pode ser exportada utilizando o ícone de descarregamento no canto superior direito.

 CNV Browser

O ícone abre um *CNV Browser* num novo separador. Este separador apresenta um gráfico que mostra a "B allele frequency" (BAF) dos SNPs na amostra analisada e dos SNPs LOH de publicações seleccionadas e da estrutura OneSeq Agilent. Um segundo gráfico representa os dados logRatio gerados pelo PureCN. Estão disponíveis mais detalhes na secção **CNV em painéis de genes em CNV Browser**.

²⁶ <https://github.com/lima1/PureCN>

13.1.5.3. CNV nos exomas e genomas

O separador de visualização CNV para dados do exoma está disponível nas análises GermlineVar, GermlineFamily, GermVar e GermVarFamily quando se utilizam ficheiros FASTQ como entrada principal e se executa uma análise CNVExome. Este separador também apresenta resultados CNV para dados do genoma em análises GermVar e GermVar Family utilizando ficheiros FASTQ, bem como em análises GermVar Tertiary e GermVar Tertiary Family utilizando ficheiros VCF ou dados CGH MicroArrays.

Metodologia para o exoma

Esta metodologia aplica-se especificamente à ferramenta CNVExome. Tal como os pipelines de análise de CNV em dados de painéis de genes, o processo de identificação de variantes neste caso depende de uma análise da profundidade de cobertura. No entanto, a análise do CNVExome utilizando a ferramenta [GATK GermlineCNVCaller](#) é individual, uma vez que cada amostra é comparada com um modelo HHMM (*Hierarchical Hidden Markov Model*) calculado pela SeqOne.

Para ter em conta as variações técnicas na preparação das amostras, este modelo baseia-se numa grande coorte que vai de uma a várias centenas de amostras sequenciadas pelo laboratório que as utiliza:

- O mesmo protocolo de preparação de bibliotecas
- O mesmo kit de captura
- A mesma plataforma de sequenciação e comprimento de read

Estes requisitos e o número variável de amostras necessárias para criar o modelo podem ser explicados pela necessidade de uma coorte que ofereça um nível de cobertura homogéneo. Devem ser tidos em conta vários factores, como a profundidade da sequenciação, o grau de variabilidade dentro da coorte e a reprodutibilidade dos protocolos de preparação das amostras.

Os metadados de género (ver **Sample metadata**) devem ser preenchidos antes de iniciar a análise CNVExome para permitir a verificação da ploidia dos cromossomas sexuais. Se não forem preenchidos, é apresentada uma mensagem de aviso "Ploidy" na interface de visualização do CNV.

Limitação: A documentação do GATK GermlineCNVCaller recomenda a exclusão das regiões PAR (PseudoAutosomal Regions), ou seja, as regiões PAR1 e PAR2, se a sua contribuição para a cobertura total do X for desproporcionada. De facto, a ploidia é difícil de calcular para as extremidades do cromossoma X correspondentes às regiões PAR para muitos modelos calculados pelo SeqOne. Como resultado, a ploidia não é calculada corretamente para estas

regiões, que são detetadas artificialmente a 1 cópia, enquanto que todo o cromossoma X é detectado a 2 cópias.

Os utilizadores são convidados a contactar a equipa de apoio ao utilizador em support@seqone.com para saberem mais sobre o procedimento de criação do CNVExome no seu grupo.

Metodologia para o genoma

Para versões até v3.6, as análises GermVar e GermVar Family em dados do genoma integram a ferramenta Sentieon CNVScope²⁷ para a deteção de CNV. A partir da versão v3.7, os worksets GermVar e GermVar Family utilizam a ferramenta Cue²⁸ para identificação de CNV em dados genómicos.

Os dados de duplicação segmentar são detectados usando o Sentieon's SegDup Caller, que é projetado especificamente para identificar eventos de número de cópias em regiões genômicas altamente homólogas em dados WGS apenas usando GermVar. Este chamador tem como alvo um conjunto curado de genes clinicamente relevantes conhecidos por serem desafiadores para as abordagens padrão de detecção de variantes. Os eventos resultantes são integrados automaticamente no pipeline de análise e visualizados no visualizador de CNVs & SVs como CNV (tanto ganho quanto perda), permitindo a identificação de variantes em regiões genômicas que antes eram inacessíveis à análise padrão.

Visualização e interpretação

Está disponível uma interface de interpretação no separador **CNV** para as análises CNVExome, GermVar e GermVarFamily, GermVar Tertiary ou GermVar Tertiary Family. A interface está dividida em duas secções: uma tabela de resultados CNV à esquerda e uma página CNV detalhada com vários separadores à direita.

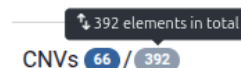
²⁷ <https://support.sentieon.com/appnotes/cnvscope/>

²⁸ <https://github.com/PopicLab/cue>

Tabela de resultados CNV

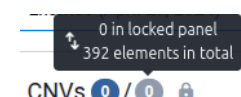
Uma tabela de resultados apresenta as CNVs identificadas, juntamente com um menu de filtros baseado na sua presença em genes OMIM mórbidos [<https://omim.org/>], tamanho, frequência de aparecimento de CNVs na coorte ou LogRatio no caso da análise CGH MicroArrays.

O número de CNVs filtrados (CNVs selecionados) e o número total de CNVs detectados (elementos no total) são apresentados na parte superior da tabela de CNVs.



392 elements in total
CNVs 66 / 392

Um ícone de cadeado indica que a análise utiliza um painel restrito, e os detalhes dos CNVs' aparecem nas dicas de ferramentas associadas.



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

É apresentada uma mensagem de aviso  Ploidy quando é detectado um valor de ploidia anormal e/ou quando os metadados relativos ao género (ver **Metadados da amostra**) não foram preenchidos antes do lançamento da análise.

Ao clicar no ícone "Ploidy", o utilizador pode visualizar a ploidia de cada cromossoma e o sexo do paciente. Os valores anómalos são realçados a amarelo.

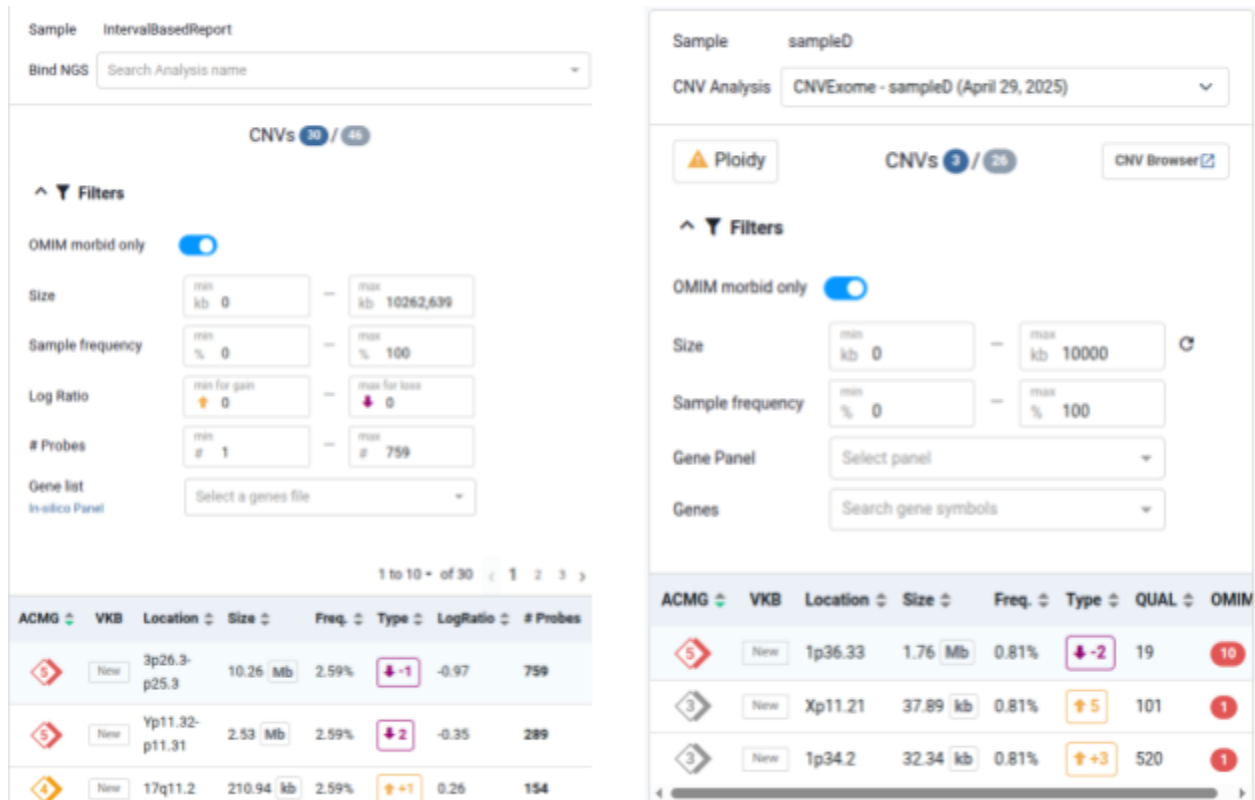


Figura 25: Overview" dos CNVs' detectados e menu de filtros de uma análise CGH (esquerda) e de uma análise padrão (direita).

No caso de uma análise CGH MicroArrays combinada com um array SNP (CGH+SNP), os eventos de loss of heterozygosity (LOH) detectados são reportados na tabela de resultados CNV. Estes eventos correspondem a perda de heterozigotia de cópia neutra (CN-LOH). Por outras palavras, a região mantém duas cópias, mas estas são idênticas, caracterizadas pela presença de SNP homozigóticos.

No caso de uma análise CGH MicroArrays, é possível selecionar uma análise SNV não-CGH (GermlineVar, GermlineFamily, GermVar Tertiary e GermVar Tertiary Family) no menu pendente "Bind NGS" para apresentar as variantes que se sobrepõem à CNV detetada.

Filtros

OMIM morbid only. Este filtro está ativado por defeito para identificar facilmente as CNVs" que afectam genes clínicos, anotados OMIM Morbid com base no seu identificador NCBI (NCBI gene_id) ou, na sua falta, com base no nome do gene e sinónimos (gene_symbol e gene_synonyms). Basta levantar este filtro para visualizar todas as CNVs' detetadas na amostra.

Size. O utilizador pode filtrar os CNVs' de acordo com o seu tamanho esperado (em kb).

Sample frequency. O utilizador pode filtrar os CNVs em função do seu tamanho e da sua frequência (Sample frequency). Esta frequência corresponde à ocorrência da CNV entre todas as amostras analisadas até à data da análise e comparadas com o mesmo modelo, incluindo as amostras utilizadas para construir o modelo. A Sample frequency baseia-se numa sobreposição de 95% e considera o mesmo tipo de CNV (gain/loss/deletion). Ao passar o rato sobre o seu valor apresentado como percentagem, o utilizador verá aparecer a fração correspondente. A frequência pode demorar vários minutos a ser actualizada após a abertura da análise.

Gene Panel. Tal como um painel in silico, este filtro permite ao utilizador seleccionar e aplicar uma lista de genes para filtrar as CNVs detetadas na análise utilizando o menu pendente (ver **Gestão dos parâmetros do utilizador**). Este filtro não está disponível se tiver sido aplicado um painel restrito à análise a partir do separador Panel (Painel) do painel lateral de gestão da análise (ver secção **Panel modal view**).

Genes. Permite-lhe filtrar os CNVs detectados que contêm os genes seleccionados através deste filtro.

Log Ratio. Para análises baseadas em dados CGH MicroArrays, o Log Ratio (entre o sinal bruto e o sinal de controlo) presente nos ficheiros de entrada da Agilent é apresentado numa coluna adicional. O filtro associado permite ao utilizador seleccionar valores mínimos para ganhos de cópias (min para gain) e valores máximos para supressões de cópias (max para loss). Os eventos LoH (Loss of Heterozygosity) continuam a ser retidos mesmo depois de este filtro ter sido aplicado.

#Probes. Para análises baseadas em dados CGH MicroArrays, a coluna #Probes indica o número de sondas utilizadas em cada região (exónica, intergénica ou intrónica). Estes valores são retirados do ficheiro de entrada IntervalBasedReport.xls, e o filtro associado permite ao utilizador seleccionar o número mínimo e máximo de sondas utilizadas. Numa análise CGH+SNP combinada, o número de sondas indica o número de SNPs homozigóticos na região LOH.

Seleção de CNV

A lista de CNVs" também pode ser ordenada por classe de patogenicidade ACMG (ACMG), posição genómica (*Location*), tamanho (*Size*), frequência de amostragem (*Freq.*), número de cópias (*Type*), pontuação de qualidade (QUAL) em GermVar Tertiary e por LogRatio em CGH MicroArrays (*LogRatio*).

ACMG". A classe de patogenicidade de cada variante é calculada semi-automaticamente de acordo com as recomendações do *American College of Medical Genetics and Genomics*. A classe de patogenicidade é o modo de ordenação predefinido na tabela CNV.

Os eventos LoH (Loss of Heterozygosity) em CGH MicroArrays são sempre atribuídos à classe 3 (VUS).

Type (Copy number). A variação do número de cópias é apresentada (por exemplo: +2, -1, +3) para CNVs' detetadas em cromossomas autossómicos (1 a 22), uma vez que o número de cópias é definido como 2 por defeito.

No entanto, é apresentado o número absoluto de cópias (ex: 0, 1, 2) para CNVs detetadas em gonossomas (cromossomas sexuais X e Y) e não a alteração fold.

Ao contrário da análise de painéis de genes, não são necessários metadados de amostras relacionados com o género para incluir cromossomas sexuais na identificação de CNV do exoma.

Para as análises CGH+SNP combinadas, os eventos CN-LOH, que retêm duas cópias, são representados na coluna Type pelo distintivo  .

QUAL. A pontuação de qualidade Phred é calculada no CNVExome pelo chamador GATK'. Quando uma CNV é composta por vários intervalos genómicos, a pontuação QUAL mais elevada entre todos os intervalos é então atribuída à CNV. Para as análises terciárias, a pontuação QUAL apresentada é a obtida a partir do ficheiro VCF de entrada (DRAGEN files ou ONT).

CNV Browser

O ícone abre um CNV Browser num novo separador que apresenta a frequência alélica B (BAF) dos SNPs (Single Nucleotide Polymorphism) presentes na amostra analisada. O gráfico é descrito na secção anterior (**CNV em painéis de genes**, em CNV Browser).

Página pormenorizada do CNV

Resumo da CNV. Este separador inicial apresenta uma grande quantidade de informações e anotações sobre o CNV selecionado, incluindo:

- Pormenores da classificação automática ACMG
- As regiões clínicas e os genes incluídos no intervalo afetado pela variante. A anotação das regiões clínicas baseia-se nas bases de dados [ClinGen](#) e [OMIM](#). Por defeito, apenas são apresentados os genes clínicos anotados pela OMIM Morbid. Para apresentar todos os genes clínicos, o utilizador pode ativar a função "Show all genes" no lado direito da página.
- Sobreposições entre a CNV detetada e a base de dados [DGV](#).
- As pequenas variantes (SNVs e indels) detetadas em genes afectados por CNV. Por defeito, o filtro SeqOne-diag está ativado. Para desativar este filtro, clique em SeqOne-diag e desmarque SeqOne-diag nos perfis de filtro. Para apresentar todas as variantes, os utilizadores podem ativar a função "Show all variants" à direita da página.
- Uma secção dedicada à base de conhecimentos sobre variantes (VKB) para identificar variantes mais de 95% idênticas que já tenham sido validadas no grupo e as respectivas Evaluations.

ACMG Classification (Automatic) ⓘ

Likely Pathogenic (+0.9)

Genomic content

Genes or Genomic regions

Gene number

Literature and databases

Inheritance patterns

1A 0

2C,1 +0.9

3A 0

No selected tags, only 40 is assessed.

Not assessed

Clinical regions

Region name	Overlap	Link	Description
CHROMOSOME 16p13.3 DELETION SYNDROME, PROXIMAL	0.29%	OMIM	
CHROMOSOME 16p13.3 DUPLICATION SYNDROME	0.29%	OMIM	
ALPHA-THALASSEMIA/IMPAIRED INTELLECTUAL DEVELOPMENT SYNDROME, DELETION TYPE	0.29%	OMIM	

Clinical genes

Show all genes

Gene symbol	Databases	Phenotype(s)	ClinGen haploinsufficiency score	pLI (GnomAD)	Exons	Fraction overlap	Variants
PKD1	omim	[AD] Polycystic kidney disease, ad...	3 (Sufficient evidence for dosage pathogenicity)	0.12051	12/46	48.39%	

DGV overlaps

4

100%

1.54%

26

11

Overlapping CNV

Max DGV overlap

Corrected frequency

Samples

Studies

ID	Fraction overlap	CNV type	CNV size	Localisation	Frequency	# Samples	# Studies
gssvG13891	100%	Gain	51.76kb	16:2,138,711-2,190,473	20.24%	17	5
gssvG13898	88.78%	Gain	25.42kb	16:2,165,630-2,191,053	0.13%	2	2
gssvL42059	44.04%	Loss	25.71kb	16:2,147,425-2,173,136	100%	2	2
gssvL42061	32.63%	Loss	10.27kb	16:2,178,475-2,188,740	7.46%	5	2

Variants

SeqOne-diag 0 / 51

Show all variants

No variant found

Figura 26: Overview" do separador "Resumo das CNV

- O IGV do navegador do genoma onde os CNVs podem ser visualizados (*não disponível nas análises GermVar Tertiary e GermVar Tertiary Family*). Para verificar a estabilidade do sinal, podem ser visualizados até cinco outros pacientes para além do indivíduo analisado a partir do menu de configuração  do gráfico que representa o número de cópias (DCN).

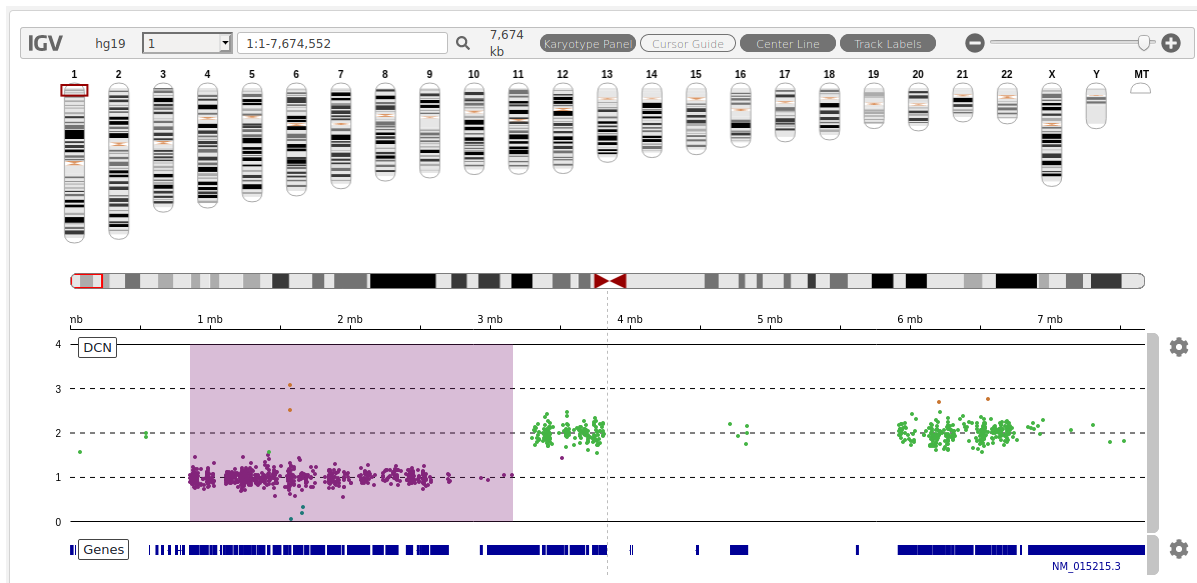


Figura 27: Vista dos CNVs' a partir de um navegador IGV integrado

O código de cores é o seguinte para o paciente selecionado: blue (0 a 0,5 cópias), purple (0,5 a 1,5 cópias), green (1,5 a 2,5 cópias), orange (2,5 a 3,5 cópias) e red (>3,5 cópias). As outras amostras são apresentadas a cinzento.

CNV Occurrences" (Ocorrências de CNV). Este separador é utilizado para identificar outras amostras que contenham a CNV, independentemente de qualquer avaliação. Para cada ocorrência, serão introduzidas as seguintes informações:

- o nome da amostra (Sample),
- o nome da análise e uma ligação para a mesma (Analysis),
- o nome do projeto (Project),
- a localização genómica da CNV (location),
- o tamanho da CNV (Size),
- o tipo de CNV,
- a sobreposição de fracções.

Os acertos são ordenados por ordem decrescente da percentagem de sobreposição. É possível filtrar os CNVs apresentados de acordo com a fração mínima de sobreposição preenchendo o campo 'min fraction overlap (%)' no topo da lista. Por defeito, este filtro está definido para 80%.

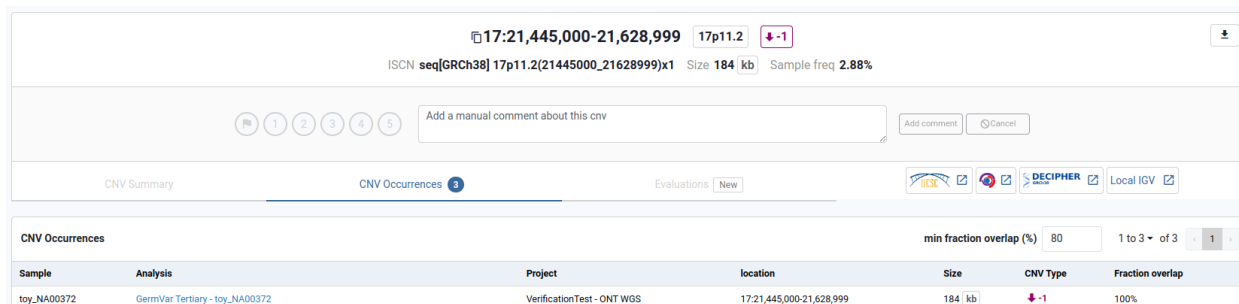


Figura 28: "Overview" do separador de ocorrências de CNV na análise CNVExome

Evaluations. A classificação de patogenicidade para cada CNV pode ser definida pelo utilizador, acompanhada de interpretações sob a forma de comentários. Cada nova avaliação também ajuda a preencher a base de conhecimentos pessoais do laboratório.

UCSC. Uma ligação externa para o [UCSC](#) Genome Browser permite aos utilizadores visualizar a região genómica da CNV. Esta página abre-se num separador adicional do navegador e inclui as definições pessoais do utilizador para esse navegador, oferecendo assim uma visualização personalizada.

CNV-Hub. Um link externo para o classificador [CNV-Hub](#) permite-lhe visualizar a classificação CNV tendo em conta várias ferramentas de classificação. A ligação abre um separador adicional do navegador e integra as definições pessoais do navegador do utilizador, oferecendo uma visualização personalizada.

DECIPHER Genome Browser. Uma ligação externa para o [DECIPHER](#) Genome Browser permite visualizar as 'CNVs' no genoma. A ligação abre um separador adicional do browser e integra as definições pessoais do browser do utilizador, oferecendo uma visualização personalizada.

Local IGV. Esta ligação redirecciona a aplicação para as coordenadas da CNV que está a ser interpretada. Para visualizar a CNV, é necessário ter carregado previamente a amostra BAM na aplicação IGV Desktop. Esta opção só é compatível com as versões do IGV Desktop ≥ 17 . IGV Desktop²⁹

Para além dos ficheiros de alinhamento, estão disponíveis vários ficheiros para descarregar a partir do separador *Files* na análise do CNVExome:

- Os resultados da análise CNV como um ficheiro tabular (cnv.tsv)
- A ploidia de cada cromossoma (contig_ploidy.tsv)
- O número previsto de cópias para cada exão no manifesto (denoise.tsv)

²⁹ <https://igv.org/doc/desktop/>.

Exportação de dados CNV



O ícone no canto superior direito permite exportar um ficheiro TSV que contém os seguintes dados para a CNV em causa:

- a sua classificação VKB" e os seus comentários,
- Classificação ISCN,
- Classificação ACMG",
- cromossoma, posições inicial e final e citobanda,
- Tamanho e tipo da CNV,
- número de cópias
- número de cópias
- sample freq
- pontuação da qualidade (escala Phred)
- Log2Ratio
- número de sondas envolvidas (Probes)
- número de genes afectados pela CNV.

As informações sobre o(s) gene(s) associado(s) são descritas nas colunas seguintes:

- nome do gene (Gene_Symbol),
- Fenótipo OMIM (OMIM_Phenotypes),
- probabilidade de intolerância genética à LoF (Pli_Score)
- Pontuações de haploinsuficiência e triplosensibilidade ClinGen (HI_Score, TS_Score)
- o número de exões sobrepostos pela CNV (Nb_overlap_exon),
- o número total de exões do gene (Nb_total_exon).

Se alguns campos descritivos disserem respeito apenas a determinados CNVs (no caso do CGH MicroArrays, por exemplo), esses campos têm o valor "NA".

13.1.6. Visualizador de variantes grandes para CNVs e SVs (LVV)

O visualizador de variantes grandes (Large Variant Viewer, LVV) permite ao usuário visualizar e analisar CNVs e variantes estruturais (SVs) em uma única aba. A interface consiste em uma tabela central de variantes, um painel de filtros à esquerda e um painel lateral que se abre à direita quando uma variante é selecionada, semelhante à aba de variantes (seção Variantes pequenas).

Em análises compatíveis, o visualizador está localizado em uma aba rotulada como "CNVs & SVs" (CNVs e SVs) ou "SVs", dependendo do tipo de dados:

- A aba CNVs & SVs exibe CNVs (Perda/Ganho), SVs (Deleção/Duplicação) e eventos LOH/UPD.
- Aba "CNV" legada: Os CNVs também são exibidos no visualizador legado para manter a compatibilidade com os fluxos de trabalho existentes.
- Os SVs são acessíveis apenas na aba "CNVs & SVs" ou "SVs".

A disponibilidade de dados no visualizador de variantes grandes depende do tipo de análise:

- WGS: Totalmente compatível (CNVs, SVs, UPDs).
- WES: Totalmente compatível (CNVs, SVs, UPDs). Os CNVs aparecem apenas se o módulo de chamada de CNV for executado como parte da análise.
- Painéis: Parcialmente compatível (apenas SVs).
- Análise terciária (VCF/aCGH):
- aCGH: Todos os eventos de número de cópias são exibidos.
- VCF: CNVs e SVs (alelos ALT simbólicos) são exibidos. SVs com sequências ALT específicas são exibidos no visualizador de variantes pequenas.

Tabela de resultados de CNVs & SVs

Semelhante à tabela de variantes pequenas, a tabela principal exibe uma linha por evento, com as seguintes colunas fornecendo métricas específicas da variante:

DiagAI. Pontuação de priorização de 1 a 100, adaptada para variantes grandes, classificando as variantes pela probabilidade de causarem doenças e serem diagnosticamente relevantes no contexto do caso (especialmente quando termos HPO são fornecidos). A pontuação DiagAI combina vários componentes, incluindo: previsão de patogenicidade através da pontuação UP², correspondência de fenótipo através do PhenoGenius (orientado por HPO) e regras de especialistas/hereditariedade e qualidade.

ACMG. O indicador de classificação ACMG baseado em variantes grandes é mostrado como uma pontuação de classificação de 1 a 5 na tabela LVV. Na gaveta de detalhes, uma aba ACMG

mostra os critérios subjacentes e um resumo de classificação ao vivo de Patogênico a Benigno (principalmente para CNV Ganho/Perda).

Classificação (ícone de bandeira) que permite ao usuário classificar a variante na VKB.

Tipo de CNV/SV. Tipo de CNV/SV (ganho ou perda), tamanho, localização e citobanda.

Regiões (exônica, intrônica, UTR) e genes sobrepostos ao CNV (nomes de genes HGNC), presença no OMIM, fração de genes codificadores e detalhes sobre éxons sobrepostos. As variantes são anotadas com genes sobrepostos e genes localizados dentro de 5 kb dos pontos de quebra (preenchimento de 5 kb).

Genes associados a uma variante são classificados por relevância clínica (por exemplo, status mórbido no OMIM, porcentagem de sobreposição, pontuações de restrição). Os três genes mais relevantes são exibidos na tabela principal. A lista completa está disponível na aba "Genes & Regions" (Genes e Regiões) do painel lateral ("Details Drawer" (Gaveta de detalhes)).

RMSK %. Porcentagem do evento sobreposto às anotações do RepeatMasker (por exemplo, regiões repetitivas).

ClinVar & Decipher (Bancos de dados clínicos). Ambas as colunas exibem o número de correspondências do banco de dados clínico correspondente. A melhor correspondência é definida como um evento onde a Sobreposição de Banco de Dados é > 50%. As classificações do ClinVar e do Decipher são mapeadas para Benigna (B), VUS ou Patogênica (P).

DGV Gold & GnomAD (Bancos de dados populacionais). A tabela exibe a frequência populacional máxima encontrada para a variante. Os eventos são considerados correspondentes se a sobreposição da variante for > 80% e o tipo de evento for equivalente. A aba Frequências no painel lateral lista todos os eventos sobrepostos com > 1% de sobreposição.

Phenotypes (Fenótipos). Ao clicar nos números exibidos nas colunas correspondentes aos resultados do algoritmo Phenogenius, o usuário pode acessar:

- a porcentagem de correspondência, que é a pontuação de correspondência mais alta entre um único gene afetado e os termos HPO do paciente.
- a fração de correspondência, que é o número de termos HPO do paciente que correspondem a pelo menos um gene afetado.
- Para análise familiar, a correspondência é realizada usando os termos HPO do probando.

SampleFreq (Frequência de amostra interna). A frequência da variante dentro do projeto e entidade atuais é exibida nas colunas correspondentes.

- As correspondências são baseadas em > 80% de sobreposição de variante e tipo de evento idêntico.

- Frequência de atualização: os dados são baseados em um instantâneo das análises realizadas antes do lançamento da análise atual. Um botão manual de "Refresh" (Atualizar) está disponível no Details Drawer para atualizar esses dados.

Colunas de pontuação. O usuário pode exibir várias pontuações:

- UP²: pontuação de previsão de patogenicidade que é um componente do DiagAI.
- pLI, LOEUF: intolerância à perda de função.
- pTriplo, pHAPLO: Triplosensibilidade e Haploinsuficiência.
- HI ClinGen, TS ClinGen.

Colunas de amostra. Estas colunas exibem métricas de eventos específicas para a fonte de dados:

- Para análises familiares (excluindo aCGH), as métricas de amostra são exibidas para todos os membros da família.
- Genotype: para CNVs onde um genótipo não é fornecido pelo chamador (caller), o sistema infere o genótipo com base no número absoluto de cópias em relação à ploidia:
 - Homozigoto (hmz) com um ponto totalmente escuro
 - Heterozigoto (htz) com um ponto meio escuro
- VAF / #Reads: frequência alélica da variante e leituras de suporte (disponível para SVs).
- #Copy: número absoluto de cópias (disponível para CNVs).
- LogRatio / Probes: disponível para dados aCGH.
- QUAL: pontuação de qualidade do chamador de variantes ou arquivo de entrada.

Classificação, Filtragem, Visualização e Interpretação

Filtragem. O painel de filtro permite ao usuário filtrar variantes com base nos dados das colunas, semelhante ao sistema de filtragem da aba Variantes pequenas (veja a seção correspondente).

- Painel de genes: filtra eventos sobrepostos a genes em uma lista de genes específica.
- Qualidade: filtra com base na coluna QUAL. Por padrão, os eventos sem pontuação de qualidade (N/A) ficam ocultos, a menos que "Keep unknown" (Manter desconhecidos) seja selecionado.
- Predefinições: as configurações de filtro podem ser salvas como predefinições para uso futuro.
- Para análises familiares, o filtro Qualidade usa a lógica "OR" (OU). A variante é exibida se pelo menos um membro da família atender ao limite de qualidade.

Classificação. Por padrão, a tabela é classificada por DiagAI Score (decrecente). O usuário pode classificar por UP², Size, Location ou Sample Freq. Entity, em ordem crescente ou decrescente.

Assim como na aba CNV herdada, o usuário pode abrir as janelas de ploidia e o navegador CNV na parte superior da tabela (veja a seção CNV: Visualização e interpretação). Um ícone de gerenciamento de colunas também está disponível na extremidade direita do cabeçalho.

Painel lateral CNV/SV (Details Drawer)

Semelhante à página de variantes (seção Variantes pequenas), selecionar uma variante no Large Variant Viewer abre uma página detalhada no lado direito da tela. Esta página contém várias abas:

Variantes. Esta aba de resumo inclui a nomenclatura da variante, a pontuação DiagAI, a lista de heterozigotos compostos suspeitos (Suspected Het Compound), informações do gene e da região, e links para bancos de dados.

ACMG. Classificação detalhada.

Genes e Regiões. a lista completa de genes e regiões associados à variante, com informações clínicas detalhadas, como status mórbido OMIM, porcentagem de sobreposição, pontuações de restrição e correspondência de fenótipo.

Classificação. Classifique a variante atual em seu VKB na parte superior do painel lateral e salve-a. A classificação atual e as ocorrências desta variante em outras análises podem ser encontradas nesta página.

Frequências. O usuário pode acessar detalhes sobre as frequências de variantes nos bancos de dados DGV Gold e GnomAD, bem como para a entidade e o projeto atual.

Bancos de dados clínicos. ClinVar & Decipher.

Viewer. O usuário pode escolher visualizar a variante no SeqOne viewer ou IGV. Na sub-aba IGV, o usuário pode ver o "Start" (Início) ou o "End" (Fim) do evento, ou "Start to End" (Do início ao fim). Na sub-aba SeqOne, o usuário pode acessar o Genome Browser e visualizar o evento em uma das várias camadas disponíveis: CNV, SV, UPD, Gene, CN (copy number), BAF, DGV.

13.1.7. Assinaturas na análise SomaCGP

Metodologia para calcular a pontuação MSI

A análise MSI é uma análise individual efectuada contra uma referência padronizada. Esta referência é constituída por um conjunto de amostras negativas previamente seleccionadas e agrupadas. A amostra do tumor é comparada com esta referência utilizando a ferramenta *MSIsensor*. É então calculada uma pontuação, representada pela percentagem de regiões instáveis de microssatélites cobertas pelo painel.

O painel de normais atualmente suportado pelo SomaCGP corresponde ao painel Agilent SureSelect Cancer CGP. A utilização de outro kit ou de um desenho personalizado derivado do CGP exigirá a criação de um painel de normais dedicado.

Metodologia para calcular a pontuação TMB

O TMB é calculado em três etapas:

1. Cálculo da dimensão das regiões de codificação incluídas no painel
2. Identificação de variantes somáticas
3. Cálculo do número de mutações somáticas por megabase (mut/Mb)

A identificação das variantes somáticas baseia-se num conjunto de filtros descritos abaixo:

- Tipo de variante e filtros de efeito
 - Apenas SNVs (MNVs são excluídos)
 - Efeitos retidos: ganho de códon de paragem, alteração do quadro de leitura, variante de emenda (aceitador e dador), ablação de transcriptos, deleção/inserção em fase, variante missense, variante sinónima.

As variantes sinónimas são excluídas do cálculo da pontuação TMB não sinónima.

- Filtros de qualidade Variant
 - Variantes detetadas por Mutect2 que passam nos filtros FilterMutectCalls
 - Profundidade de cobertura (DP) $\geq 50X$
 - Índice de qualidade > 30
 - Rácio de suporte alélico (ASR) $> 0,1$
- Filtros de variantes somáticas
 - Frequência do alelo variante (VAF) entre 5% e 40%.
 - Frequência na população geral (GnomAD) $\leq 0,0001$

Riscos da utilização de amostras de controlo de TMB sintético

- Distribuição de mutações não naturais:
 - As mutações artificiais podem não refletir o contexto genómico dos tumores reais (por exemplo, agrupamento de mutações, polarização de cadeias, contexto de sequência).
- Ausência de heterogeneidade tumoral:
 - As amostras sintéticas não imitam a diversidade clonal ou a pureza do tumor, o que afecta a distribuição da frequência de alelos variantes (VAF).
 - Sem ruído de fundo ou artefactos FFPE: as amostras reais têm artefactos de sequenciação, danos FFPE e variações do número de cópias, que estão ausentes nos materiais sintéticos.
- Diferentes tamanhos de fragmentos e comportamento químico:
 - As construções sintéticas comportam-se frequentemente de forma diferente durante a preparação da biblioteca, a captura híbrida e a sequenciação.

Visualização e interpretação

O separador "*Signatures*" apresenta os resultados das análises MSI e TMB em 2 secções:

Microsatellite Instability. Esta vista mostra informações como :

- o número de regiões de microssatélites com uma cobertura maior ou igual a 50X (Sites above 50X coverage'),
- o número de sítios somáticos,
- Pontuação MSI em percentagem e intervalo de confiança (IC) de 95%,
- Status" do MSI

Tumor Mutational Burden. Esta vista mostra as seguintes informações:

- a dimensão da região de codificação expressa em Mb (Coding region size'),
- o número de variantes somáticas e não sinónimas que passam nos filtros (Variantes somáticas que passam nos filtros e Variantes somáticas não sinónimas que passam nos filtros),
- a pontuação TMB de todas as variantes somáticas que passam nos filtros em mut/Mb (TMB Score),
- a pontuação TMB das variantes não sinónimas que passam nos filtros mut/Mb.

Cada resultado pode ser descarregado como ficheiro tsv clicando no ícone  no canto superior direito de cada secção.

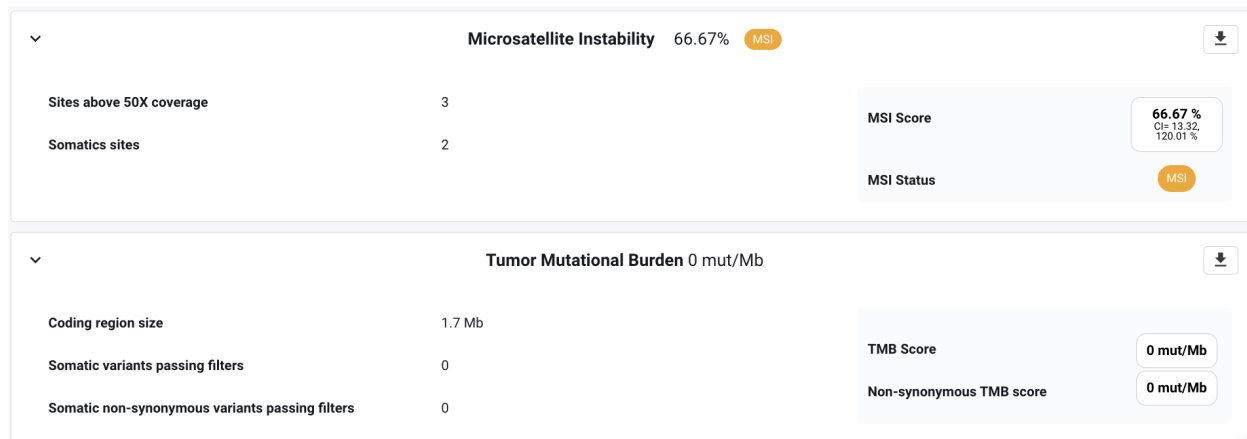


Figura 29: Visualização do separador “Signatures” de uma análise SomaCGP.

13.1.8. Repetições curtas em tandem (STRs)

A visualização de Short Tandem Repeats (STRs) está disponível para os "worksets" de análise GermVar Tertiary e GermVar Tertiary Family dedicados à análise de dados de Sequenciação do Genoma Completo (WGS), quando gerados a partir de sequenciação da Oxford Nanopore Technologies.

A detecção de repetições curtas em tandem (STR) também está disponível nos conjuntos de trabalho de análise GermVar e GermVar Family para WGS a partir de dados FASTQ gerados com Illumina, Element Biosciences, MGI e Oxford Nanopore Technologies.

Metodologia

Repetições curtas em tandem a partir de FASTQ

Os dados de repetição curta em tandem (STR) são detectados usando o ExpansionHunter, que processa o arquivo BAM juntamente com um genoma de referência e um catálogo de variantes definindo os loci alvo. O VCF resultante é então anotado pelo Stranger, que classifica cada locus como Normal, Pré-mutação ou Patogênico com base na contagem de repetições observada e nos limites de doenças conhecidos.

Ambas as ferramentas são integradas diretamente nos pipelines GermVar e GermVar Family.

ExpansionHunter: <https://github.com/Illumina/ExpansionHunter>

Stranger: <https://github.com/moonso/stranger>

Repetições curtas em tandem a partir de VCF

Os dados de repetições em tandem são extraídos do ficheiro VCF (sample.wf_str.vcf.gz) gerado a partir do workflow wf-human-variation v2.3.0 da aplicação EPI2ME (Oxford Nanopore Technologies) e fornecido pelo utilizador. Uma vez que estes dados foram anotados para cada locus portador de STR, a partir de um catálogo de variações mantido pelo EPI2ME, o nosso pipeline não modifica as anotações. O estado pré-mutado ou patogénico de cada alelo é determinado de acordo com o número de repetições sequenciadas e os limiares conhecidos de pré-mutação e patogenicidade.

Visualização e interpretação

O separador STRs permite visualizar os Short Tandem Repeats sob a forma de uma tabela que apresenta os seguintes detalhes para cada locus genotipado:

Informações e anotações sobre o Locus

- **Gene**: nome do gene (HGNC) e ligação correspondente.
- **Site**: as coordenadas genómicas da repetição.
- **Repeat Unit**: o motivo nucleotídico repetido.

- **Disease:** a doença associada de acordo com a base de dados OMIM e a ligação correspondente.
- **Transmission:** o modo de transmissão conhecido.
- **Premutation range:** o intervalo de pré-mutação correspondente ao número mínimo e máximo de repetições que determinam o estado de pré-mutação.
- **Pathogenic range:** o limiar inferior de repetições que determinam o estatuto patogénico.

Pormenores específicos de cada amostra

- **GT:** o genótipo homozigótico, bialélico ou heterozigótico do Locus.
- **Allele repeats:** o número de repetições por alelo. Os loci homozigóticos têm um único valor. Para loci bialélicos e heterozigóticos, é indicado o número de repetições para ambos os alelos. O estatuto de cada alelo é determinado pelo número de repetições: tipo selvagem a verde, pré-mutado a laranja e patogénico a vermelho.
- **Supp. reads:** o número de reads de suporte para cada alelo. Os loci homozigóticos têm um único valor, enquanto os valores para 2 alelos são apresentados para os loci bialélicos e heterozigóticos.

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: Visualização do separador STRs de uma análise GermVar Tertiary.

O quadro pode ser descarregado como ficheiro repeats.tsv através do ícone  no canto superior direito do quadro. Este ficheiro também está disponível no separador "Files" da análise.

13.1.9. Fusão de Genes

13.1.9.1. Fusões detetadas a partir do ADN

Metodologia

O módulo [GRIDSS](#) (Genome Rearrangement IDentification Software Suite), especializado na detecção de rearranjos genómicos e implementado nos worksets SomaVar e SomaVar LF, é utilizado para detetar fusões de genes a partir de sequências de ADN, enriquecidas por captura dirigida a um painel de genes

Visualização e interpretação

O separador "Fusions" de uma análise somática (Somatic LF, Somatic LF, SomaCGP, SomaVar Tertiary) apresenta a lista de fusões que passam nos vários filtros GRIDSS sob a forma de uma tabela que detalha informações sobre a fusão e cada parceiro envolvido:

Informações sobre a fusão

- fusão com os símbolos genéticos de cada parceiro,
- o tipo de evento (translocação, duplicação, inversão ou supressão),
- a ocorrência da fusão na base de dados COSMIC,
- a frequência do alelo (VAF),
- o número de reads que apoiam a fusão,
- cobertura total,
- filtros atribuídos pelo GRIDSS.

Informações relativas aos parceiros

- as coordenadas genómicas do ponto de fusão (breakend),
- o transcript RefSeq em causa,
- o sítio correspondente ao ponto de fusão (breakpoint): 5'UTR, 3'UTR, Intron, Exon (ponto de quebra não codificante), CDS (ponto de quebra codificante) ou sítio de splicing,
- a numeração do exão ou do intrão envolvido na fusão, em relação ao número total de exões ou intrões.

Um menu de filtros permite-lhe

- procurar uma fusão específica ou um gene envolvido numa fusão através de uma barra de pesquisa,
- mostrar as fusões que não passaram nos filtros GRIDSS (LOW_QUAL) marcando a caixa 'Show filtered fusions'. Por defeito, esta caixa não está seleccionada.

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: Visualização do separador Fusões para uma análise SomaVar ou SomaVar LF.

Os dados da tabela podem ser ordenados de acordo com as coordenadas genómicas (breakend), a presença da fusão no COSMIC (Databases), a frequência alélica (VAF) e o número de reads que suportam a fusão (Supporting reads).

A ligação associada às coordenadas genómicas de cada ponto de quebra abre uma vista dividida do ponto de fusão no IGV. As dicas de ferramentas detalham o número de split reads e pares não correspondentes na coluna Supporting reads, e a cobertura em cada ponto de rutura (Cov. 1 e Cov. 2) na coluna Coverage.

A coluna Filtro apresenta 3 filtros de qualidade utilizados pelo GRIDSS e que dão uma indicação da confiança na chamada de fusão:

- PASS': a fusão foi aprovada nos filtros de qualidade. A pontuação de qualidade é superior ou igual a 500 e é considerada uma chamada de confiança média ou elevada (se > 1000).
- LOW_QUAL: a fusão tem um índice de qualidade inferior a 500 para pelo menos um dos dois breakends envolvidos na fusão e é considerada uma chamada de baixa confiança'.
- ASSEMBLY_ONLY: a fusão é chamada com base numa montagem e não com base na read que a suporta. A montagem refere-se à reconstrução de sequências em torno de potenciais pontos de fusão, que são depois realinhadas para identificar a fusão.

Se estiverem disponíveis vários filtros para a mesma fusão, um sinal "+" apresenta todos os filtros na coluna Filtro.

Para além dos resultados apresentados na interface, um ficheiro fusions.tsv que detalha as

fusões detetadas está disponível para transferência no separador "Files" ou no ícone  no canto superior direito da página. Este ficheiro contém as seguintes informações:

- nome da fusão com símbolos de genes para cada parceiro (fusão)
- coordenadas genómicas de cada ponto de fusão (breakend_1 e breakend_2)

- transcrição canónica para cada parceiro (transcript_canonical_1 e 2)
- sítio correspondente ao ponto de fusão para cada parceiro (sítio_1 e sítio_2)
- Numeração do exão ou do intrão de acordo com o transcript RefSeq associado (exon_1 e 2)
- 60 pb de sequências de flanco que rodeiam o ponto de fusão (flanking_region_1 e 2)
- filtros de qualidade atribuídos pelo GRIDSS (filtro)
- tipo de evento (type)
- número de split reads (split_reads)
- número de pares de leituras discordantes (discordant_pairs)
- cobertura para cada parceiro (coverage_1 e coverage_2)
- fração de reads que suportam a fusão (read_fraction)
- número de reads que apoiam a fusão (supporting_reads)
- cobertura total (cobertura)
- Identificador de fusão COSMIC (cosmic_id)

Este ficheiro contém:

- Fusões visualizadas na interface, tendo passado os filtros de qualidade (filtro = PASS),
- Fusões detetadas mas com um índice de qualidade inferior a 500 para pelo menos um dos dois breakends envolvidos na fusão (filter = LOW_QUAL).

O índice de qualidade de cada breakend encontra-se no ficheiro gridss.raw.vcf.gz (coluna QUAL), que pode ser descarregado do separador Files tab.

13.1.9.2. Fusões detetadas a partir de ARN

Ferramenta SomaRNA para a análise de dados RNA-seq provenientes de abordagens orientadas, como a captura e o amplicon, o que significa que os utilizadores podem identificar e visualizar eventos de fusão de genes ao nível dos 'Transcripts'.

Metodologia

É utilizada uma combinação de duas ferramentas para detetar e visualizar as fusões por SomaRNA:

- [STAR](#): um alinhador de dados RNA-seq concebido para alinhar sequências não contíguas diretamente com o genoma de referência
- [Arriba](#): detecta fusões de genes a partir de alinhamentos quiméricos

Com base nos alinhamentos quiméricos do STAR, o Arriba aplica um conjunto de filtros para eliminar os artefactos conhecidos e os "Transcripts" observados em contextos não patológicos. Atribui uma pontuação de confiança a cada potencial evento de fusão que depende de vários critérios:

- O número de sequências que suportam a fusão
- O equilíbrio entre split reads e companheiros discordantes
- A distância entre os pontos de quebra, a sua proximidade (intragénica ou não)
- O tipo de evento

As fusões detetadas com Média ou Alta Confiança pelo Arriba são apresentadas no SeqOne. As fusões com Confiança Baixa são listadas nos ficheiros "arriba_discarded" (ver abaixo). A confiança do Arriba reflecte a probabilidade de um transcrito aberrante não observado em tecido saudável, a possibilidade de um rearranjo genómico subjacente e a exclusão de artefactos.

Mais pormenores sobre o nível de confiança na documentação do Arriba: <https://github.com/suhrig/arriba/blob/master/documentation/08-Interpretation-of-results.md>

Para compreender por que razão uma fusão é "Low confidence", consultar a coluna 'filters' (AA) no ficheiro "discarded". Esta indica o filtro que descartou o evento. Pormenores de cada filtro: <https://github.com/suhrig/arriba/blob/master/documentation/11-Internal-algorithm.md>

Visualização e interpretação

A partir do separador Fusões de uma análise de SomaRNA, o utilizador obtém uma lista de previsões de fusão, sob a forma de diferentes separadores.

A parte superior da página apresenta pormenores sobre a fusão e cada um dos parceiros envolvidos:

Informações gerais sobre a fusão

- o número de reads da junção de spanning, igual à soma das respectivas reads de splits de cada um dos genes 5' e 3' envolvidos na fusão,
- o número de companheiros discordantes,
- o índice de confiança atribuído pelo Arriba (High, Medium ou Low),
- o tipo de evento que conduz à fusão (translocação, duplicação, inversão ou deleção),
- o impacto da fusão no quadro de leitura,
- se aplicável, a ocorrência da fusão na base de dados COSMIC.

Informações sobre cada parceiro de fusão

- o símbolo de cada gene,
- o Refseq transcript ou, na sua falta, o ENSEMBL para cada gene,
- a numeração dos exões envolvidos na fusão,
- as coordenadas genómicas de cada ponto de quebra,
- a cobertura total no ponto de fusão (Coverage),

- a fração read que apoia a fusão.

Split reads são reads alinhadas predominantemente no gene 1 ou no gene 2 e que se sobrepõem à fusão. No caso de pares de reads discordantes, uma das reads está no gene 1 e a outra no gene 2, sem necessariamente se sobrepor à fusão.

A fração de leitura é a soma das leituras divididas e dos pares de leituras discordantes associados a um gene, dividida pela cobertura total. O valor da fração de leitura pode ser superior a 100% quando as leituras split read associadas a um gene também fazem parte de pares de leituras discordantes. Estas reads são então contadas duas vezes em vez de uma.

DHRX - TCEANC
E3 - E4

POLR2J2 - EFCAB13
E3 - E18

POLR2J2 - EFCAB13
E4 - E18

MEGF8 - CNFN
E3 - E4

General informations

of spanning junction reads

568

of discordant mates

246

Confidence ⓘ

high

Fusion type

translocation

Reading frame

in-frame

Database

cosmic

Partners informations

5prim

3prim

Gene

NPM1

ALK

Transcript

ENST00000517671 ⓘ

NM_004304.4

Exons

5 / 12

20 / 29

Breakpoint

5 : 170818803 (+)

2 : 29446394 (-)

Coverage

14004

8455

Read fraction

3.9%

6.08%

Figura 32: Visualização da informação de fusão de uma análise de SomaRNA.

A parte inferior da página mostra representações gráficas da fusão, tanto ao nível do gene e do exão, como ao nível da proteína. Esta última secção está dividida em dois separadores:

- O primeiro separador apresenta a estrutura da fusão e os parceiros envolvidos, onde a cobertura dos exões abrangidos pelo manifesto de análise é representada como um diagrama sobre um fundo amarelo:

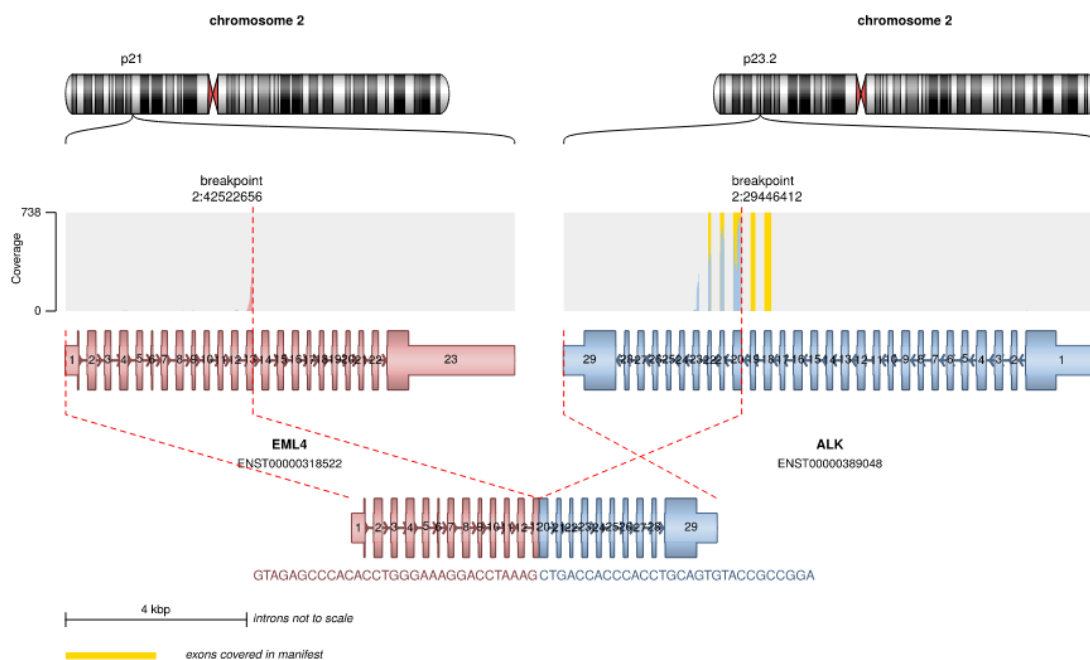


Figura 33: Vista de um evento de fusão entre dois genes

- Os domínios proteicos retidos após o evento de fusão são destacados numa segunda vista:

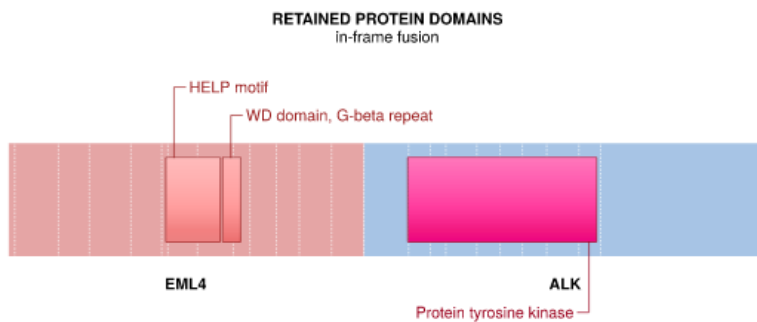


Figura 34: Vista do polipéptido resultante de uma fusão entre dois Genes

Para além dos resultados apresentados na interface, estão disponíveis vários ficheiros para descarregar:

- As fusões retidas pelo Arriba em formato TSV (ficheiro arriba.tsv),
- As fusões eliminadas pelos filtros Arriba em formato TSV (ficheiro arriba_discarded_standard.tsv),
- as fusões eliminadas pelos filtros Arriba e listadas no COSMIC em formato TSV (ficheiro arriba_discarded_cosmic.tsv),
- as diferentes figuras apresentadas na interface em formato SVG

13.1.10. Clonalidade

O «IGHV mutational status report» no separador «Clonality» da análise SomaHemato mostra os resultados do estado mutacional.

A secção «IGHV Results» apresenta:

- Estado mutacional IGHV e percentagem de homologia:
 - Mutated (Mutado): se a homologia for <98% de identidade com a linha germinativa
 - Unmutated (Não mutad): se a homologia for ≥98% de identidade com a linha germinativa
 - Borderline (Limítrofe): para clones na faixa de identidade de 97–98%
 - Polyclonal (Policlonal): nenhum clone principal acima de 20%
 - Non-Conclusive (Não conclusivo): se qualquer QC essencial falhar
- O subconjunto determinado (com impacto clínico) e confiança: *2 (alto risco), *8 ou Nenhum
- Uma notificação de aviso se houver avisos no relatório.

A secção «Clonality» lista os clones identificados com as suas frequências e estado de produtividade (produtivo, não produtivo).

Cada clone tem uma secção dedicada que mostra o seu estado mutacional, frequência, link para a página IMGT_quest e detalhes adicionais (homologia, completude), juntamente com verificações de controlo de qualidade.

13.1.11. Variantes de splicing

Metodologia

O workset SomaRNA utiliza o conjunto de ferramentas *Regtools* para detetar saltos de exões e identificar variantes de emenda alternativas a partir de sequenciação de RNA, enriquecida por captura dirigida a um painel de genes. O RegTools extrai as junções exão-exão do ficheiro BAM gerado pelo alinhador de dados STAR RNA-seq e anota-as de acordo com uma estrutura de transcrição conhecida com base num ficheiro GTF e nas anotações Ensembl.

Visualização e interpretação

O separador Variantes de Splice de uma análise SomaRNA está focado na visualização das variantes de splice dos seguintes genes clinicamente relevantes e da sua principal variante acionável:

- MET (MET Proto-Oncogene, Recetor Tyrosine Kinase) e a variante acionável MET Δ 14
- EGFR (Recetor do Fator de Crescimento Epidérmico) e a variante de splicing EGFRvIII
- AR (Recetor de Androgénio) e a variante de splicing acionável AR-V7

Uma lista completa das variantes de emenda detetadas na análise de SomaRNA, para além das enumeradas acima, pode ser descarregada diretamente como um ficheiro TSV através do ícone Download na interface ou do separador Files na análise.

Para cada gene mencionado acima, é apresentado o estado de deteção (Detectado ou Não detectado) da principal variante acionável, juntamente com o número de outros eventos alternativos detectados para esse gene.

Os utilizadores podem ver detalhes das principais variantes de emenda detetadas abrindo uma inserção dedicada, que está dividida em três partes:

- À esquerda, são apresentadas informações sobre a principal variante de splice acionável: coordenadas genómicas, nome do gene, RefSeq transcript, tipo de variante detetada detalhando os exões em falta, fase de leitura, fração de reads e número de reads que suportam a variante. Finalmente, é especificada a média de reads de suporte canónico. Isto corresponde à média das reads 5' e 3' que suportam a inclusão do exão.
- à direita, um gráfico sashimi (Junction Plot) ilustra as junções detetadas entre exões, bem como o número de reads que suportam cada junção.
- na parte inferior, são listadas as informações sobre os eventos alternativos detectados, ou seja, isoformas diferentes da variante principal de interesse descrita acima: tipo de variante detetada (números de exões em falta entre parênteses), fase de leitura, fração e

número de leituras que suportam a variante, número médio de leituras que suportam o transcript canónico (Avg. canonical support), status PASS ou FAILED de acordo com os filtros descritos abaixo.

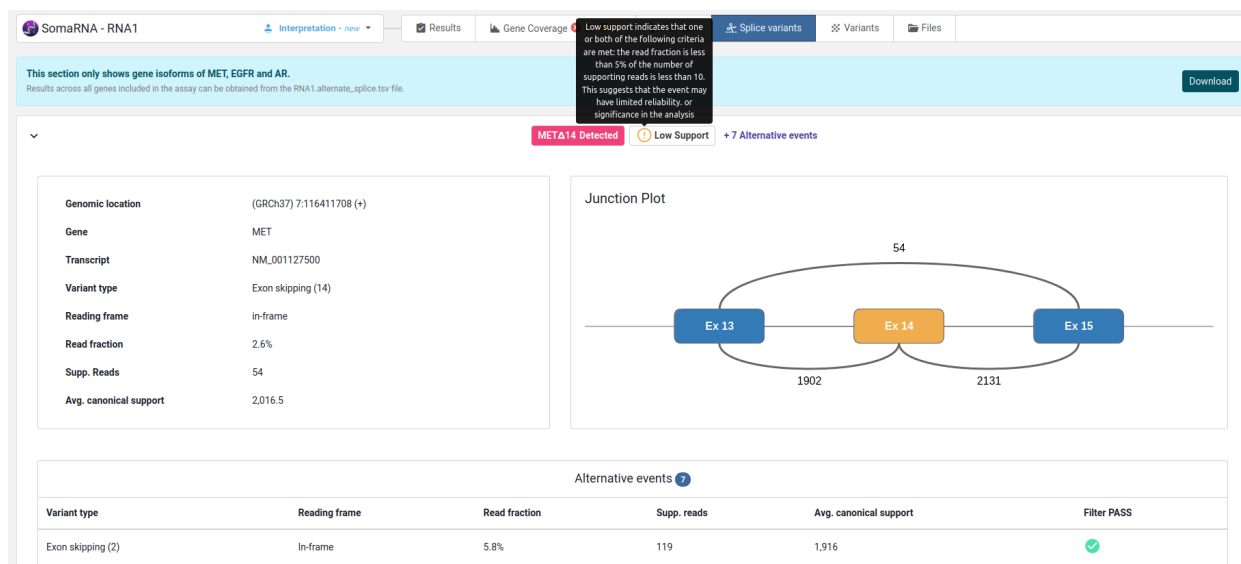


Figura 35: Visualização do separador Splice variants de uma análise SomaRNA.

A deteção de variantes de emenda nos genes apresentados no separador Variantes de emenda não está sujeita à aplicação de filtros de deteção.

No entanto, um ícone "Low support" associado à deteção de METΔ14 e EGFRvIII indica ao utilizador que as métricas relacionadas com as reads de suporte estão abaixo dos seguintes limiares: fração de reads < 5% ou número de reads de suporte < 10.

O estado PASS' ou FAILED de outros eventos alternativos também utiliza estes limiares:

- **PASS** : fração de reads $\geq 5\%$ e número de reads ≥ 10 .
- **FAILED** : fração de reads < 5% ou número de reads < 10.

O ficheiro TSV descarregável detalha as seguintes informações para todos os eventos de splicing detectados: RefSeq transcript", gene, direção da cadeia (+ ou -), fração de leituras e número de leituras que suportam a variante, número de exões em falta e números de exões correspondentes, fase de leitura, coordenadas genómicas das junções. O ficheiro também especifica, em FILTER_PASS (yes/no), se a variante passa os limiares aplicados às reads de suporte (fração de reads $\geq 5\%$ e número de reads ≥ 10) e, em Actionable (yes/no), a possibilidade de ação da variante.

13.1.12. Estado MSI/MSS

O workset de análise SomaMSI está disponível quando se executa uma nova análise nos dados do painel de genes e é utilizado para determinar o estado MSI ou MSS de uma amostra de tumor.

Metodologia

Esta análise de coortes não individuais é semelhante a uma análise CNVCapture ou SomaCNVCapture". São necessárias pelo menos oito amostras para efetuar a análise.

A ferramenta gera uma amostra de referência a partir das amostras da coorte que se estima serem negativas. Uma vez que este estado das amostras não é conhecido antecipadamente, esta estimativa baseia-se no cálculo da taxa de deleções observadas nas regiões de microssatélites: metade das amostras da coorte com a taxa mais baixa são assim selecionadas para gerar uma amostra de referência, contribuindo cada amostra com uma fração das suas sequências.

Cada amostra de tumor é subsequentemente comparada com a amostra de referência utilizando a ferramenta MSIsensor'. É calculada uma pontuação, que representa a percentagem de regiões de microssatélites estudadas que apresentam instabilidade. As regiões estudadas são as regiões de microssatélites abrangidas pelo painel e que são detetadas pelo MSIsensor'.

Com base nesta pontuação, o estado MSI ou MSS das diferentes amostras é ditado por um limiar que é calculado da seguinte forma: $\max(10, \text{median} + \max(3 * \text{median-abs-dev}, 0.3 * \text{median}))$

O princípio subjacente a esta pontuação é o seguinte:

- Numa análise que apresente amostras MSI e MSS, o limiar será geralmente a mediana + $3 * \text{median-abs-dev}$.
- No entanto, se todas as amostras tiverem o estado MSS, se as pontuações forem muito baixas e se a sua distribuição for extremamente próxima, o limiar será fixado em 10 para evitar gerar falsos positivos.

Visualização e interpretação

Uma vez terminada a análise, os resultados de todas as amostras podem ser visualizados a partir da análise SomaMSI. O separador Resultados está dividido em duas secções:

MSI graph: representação gráfica dos resultados em que cada amostra é apresentada de acordo com a sua pontuação. O limiar determinado pela coorte de amostras cria duas zonas no gráfico que representam as amostras consideradas MSI e MSS, respetivamente.

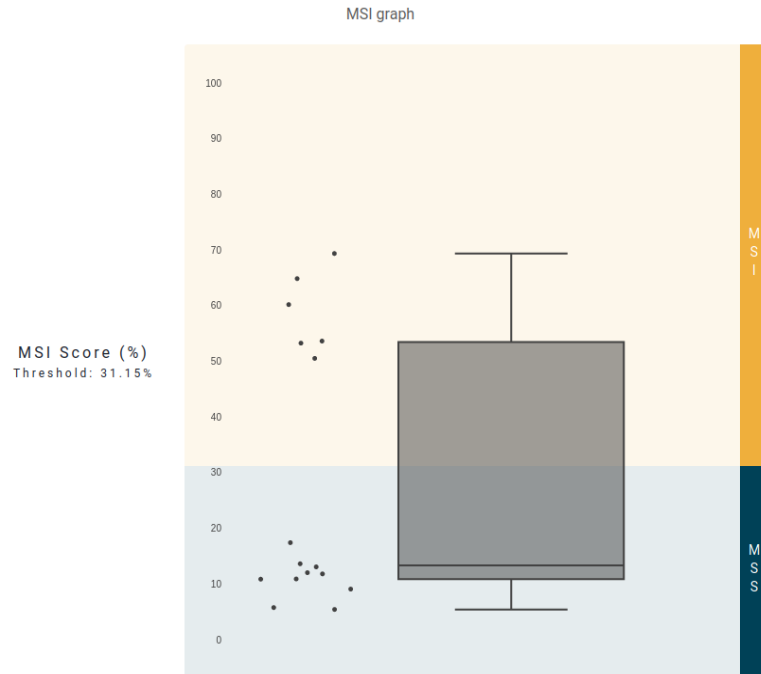


Figura 36: Box plot representando a distribuição dos scores MSI para a coorte da amostra.

Per sample details: a repartição das pontuações e o estado atribuído a cada amostra estão disponíveis no painel da direita.

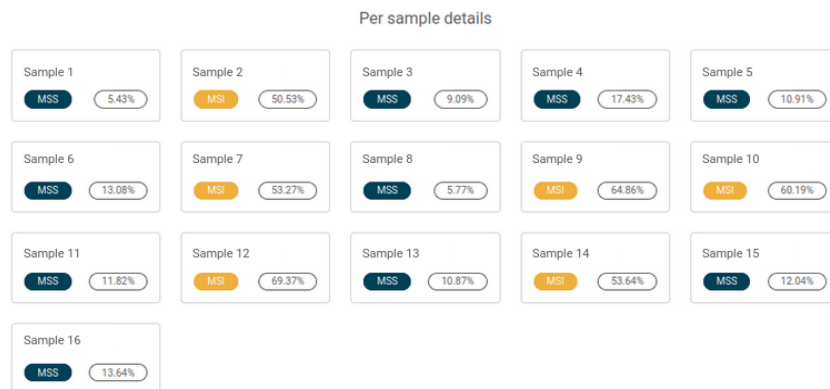


Figura 37: Repartição das pontuações e do estado de todas as amostras incluídas na análise do SomaMSI.

13.1.13. Estado HRD/HRP

Metodologia

A detecção do estado de HRD proposta pela SeqOne baseia-se na análise de dados de sequenciação do genoma de "Low coverage" (sWGS), aos quais podem ser adicionados os resultados de uma análise de um painel de genes, se este incluir os genes BRCA1 e BRCA2.

O princípio é o seguinte:

- Quantificar a instabilidade genómica a partir de dados sWGS e identificar grandes rearranjos (LGA), bem como perdas de cópias parentais (LPC), o que reflectiria uma perda de heterozigotia.
- Identificar variações do número de cópias (CNVs) em genes-chave descritos como potencialmente afectando o estado HRD.
- Detetar opcionalmente variantes patogénicas nos genes BRCA1, BRCA2 e TP53, se os resultados do painel estiverem disponíveis na interface SeqOne para o paciente correspondente. Neste caso, os tipos de variantes detectados pelo workset SomaVar que são tidos em conta são SNV, indels e SV (detectados por 'Freebayes', 'Mutect2' e GRIDSS).

Status HRD. O estado HRD positivo é definido por pelo menos um dos seguintes factores (**apenas nas versões SomaHRD v1.2 e v1.3**): (1) pelo menos uma mutação BRCA1/2 patogénica e/ou (2) uma HRD probability maior ou igual a 0,5.

A partir da versão v1.4 do workset SomaHRD, o estado da análise é definido unicamente pela pontuação de instabilidade genómica (pontuação GI). Os utilizadores podem seleccionar um HRD status (Positive, Negative, ou Non-conclusive) ao lado do GI status para o incluir no relatório em vez do GI status.

HRD probability. (**Apenas versões SomaHRD v1.2 e v1.3**) A SeqOne HRD Probability corresponde à probabilidade de o tumor ser HRD positivo com base nas características de instabilidade genómica LGA e LPC determinadas a partir dos dados sWGS. As Grandes Alterações Genómicas (LGA) são definidas como o número de cópias de pontos de quebra, em que um ponto de quebra corresponde a uma alteração no número de cópias entre dois segmentos genómicos com, pelo menos, 10 Mb de comprimento e não mais de 3 Mb de distância. As Losses of Parental Copy (LPC) são definidas como o número de segmentos haplóides com pelo menos 10 Mb de comprimento. A probabilidade de HRD é uma combinação linear das duas características de instabilidade genómica, convertida através de regressão logística para colocar os valores numa escala de probabilidade entre 0 e 1.

GI probability (GiP)(**do SomaHRD v1.4**) A probabilidade de instabilidade genómica (GI) representa a probabilidade de o tumor apresentar instabilidade genómica, com base na

'Shallow Whole Genome Sequencing' (sWGS). O cálculo desta probabilidade depende dos dois tipos de características de instabilidade genômica LGA e LPC. A LGA é definida como o número de pontos de quebra do número de cópias. Um ponto de quebra corresponde a uma alteração do número de cópias entre dois segmentos genômicos suficientemente grandes e muito próximos. A pontuação LGA não é uma medida única, mas sim um conjunto de diversas variações. Estas variações são calculadas utilizando diferentes parâmetros para o tamanho mínimo do fragmento e a distância entre fragmentos. A pontuação LPC, por outro lado, refere-se ao número de segmentos haplóides com pelo menos 10 Mb de comprimento. A probabilidade de instabilidade genômica SeqOne não é uma função direta da média das pontuações LGA e LPC. É calculada utilizando um modelo de reforço que integra várias características LGA e LPC, melhorando assim a exatidão da previsão. A probabilidade final é a média de 100 probabilidades (variando de 0 a 1) calculadas pela aplicação de diferentes parâmetros do modelo.

GI status. (de SomaHRD v1.4) Um estado de instabilidade genômica positivo (GI status) é definido por uma probabilidade de instabilidade genômica superior a 50%.

BRCA Status. Para a detecção de variantes patogênicas, determinando o estado BRCA independentemente da probabilidade HRD ou da probabilidade GI, são aplicados os seguintes critérios às variantes BRCA1 e BRCA2 da análise SomaVar do paciente:

- Allele frequency $\geq 15\%$ (**versões SomaHRD v1.2 e v1.3**),
- Allele frequency $\geq 10\%$ (**de SomaHRD v1.4**)
- Coverage ≥ 35
- Qualidade média de cada base portadora da variante (pontuação Phred) ≥ 20
- Observação alternativa (número de sequências que transportam a mutação) ≥ 5
- Classificação 4 (*provavelmente patogénica*) ou 5 (*patogénica*) de acordo com a nossa ferramenta de classificação da patogenicidade das variantes baseada nos critérios ACMG.

TP53 variants. Os seguintes critérios são aplicados às variantes patogênicas TP53 da análise SomaVar realizada no paciente:

- Coverage ≥ 100
- Qualidade média de cada base portadora da variante (pontuação Phred) ≥ 20
- Observação alternativa (número de sequências que transportam a mutação) ≥ 5
- Classificação 4 (*provavelmente patogénica*) ou 5 (*patogénica*) de acordo com a nossa ferramenta de classificação da patogenicidade das variantes baseada nos critérios ACMG.

Estas variantes são apenas apresentadas sem um limiar mínimo de frequência alélica no relatório de análise SomaHRD, mas não são determinantes do estado HRD. Apenas a variante patogénica TP53 com a frequência alélica mais elevada será indicada no relatório HRD.

Cellularity. Os metadados de celularidade (% de células tumorais na amostra) são necessários para executar a análise SomaHRD. Pode ser introduzido quando a amostra é carregada ou antes do lançamento da análise, editando os metadados da amostra correspondente (ver **Sample metadata**). Se o valor for inferior a 20%, o resultado será não-conclusivo. No entanto, o estado HRD pode ser positivo se a análise for acompanhada por um painel que inclua BRCA1 e BRCA2, o que dá um estado BRCA positivo.

Procedimento de aviso de fração tumoral: A análise SomaHRD inclui um mecanismo de alerta para indicar quando uma amostra pode ter um conteúdo tumoral baixo. Este mecanismo baseia-se numa medida estatística denominada "overlap statistic" (estatística de sobreposição), que avalia a semelhança entre segmentos genómicos consecutivos em termos de rácio do número de cópias. As amostras com sobreposição significativa têm frequentemente fracções tumorais mais baixas, o que torna mais difícil distinguir a instabilidade genómica.

Para estabelecer um limiar fiável para este aviso, correlacionámos a estatística de sobreposição com a Variant Allele Frequency (VAF) das mutações TP53 utilizando dados clínicos. É acionado um alerta se a estatística de sobreposição exceder um limiar predefinido. Este alerta destaca potenciais discrepâncias com a percentagem de conteúdo tumoral fornecida manualmente, o que pode indicar que a fração tumoral é inferior a 20%. As amostras com fracções de tumor abaixo deste nível correm um risco elevado de produzir falsos negativos. No entanto, essas amostras não são rejeitadas e o alerta serve de orientação para uma interpretação cuidadosa dos resultados de HRD.

Minimum coverage. A cobertura mínima necessária para efetuar o teste é 0,1X, mas recomendamos uma cobertura mínima de 0,5X, idealmente 1X. Quando a cobertura é reduzida para 0,1X, a taxa de erro de classificação aumenta em 5%.

Embora os dados sWGS possam ter sido gerados a partir de um kit que tenha UMIs, estes não são tidos em conta durante a análise SomaHRD. O workset inclui um passo de corte para adaptadores e UMIs.

Visualização e interpretação

Uma vez terminada a análise, os resultados estão disponíveis a partir da análise SomaHRD. Para além das métricas QC e do relatório técnico, um relatório HRD é diretamente acessível a partir do separador HRD Report (**versões SomaHRD v1.2 e v1.3**).

A partir da **versão v1.4 do worksets SomaHRD**, os resultados da análise são apresentados num separador HRD dedicado. Cada secção do separador reflecte uma secção do relatório HRD descrito abaixo (HRD Summary, Sample quality, Genomic instability features, BRCA1/BRCA2 & TP53 mutations, Copy number variations).

A partir deste separador, o utilizador pode exportar estes resultados como um ficheiro CSV ou como um relatório HRD em formato PDF. No caso de uma exportação em PDF, o utilizador pode seleccionar o idioma do relatório HRD a partir de uma lista de idiomas disponíveis: Francês, Espanhol, Inglês, Italiano, Alemão. A seleção predefinida é o inglês.

O relatório HRD está dividido nas seguintes secções

Um **header** lista os metadados do paciente e da análise, o tipo de teste efectuado e o HRD status (**SomaHRD v1.2 & v1.3**) ou o GI Status (**SomaHRD ≥ v1.4**).

Sample quality. Esta secção resume as principais métricas de qualidade da análise, bem como os limiares mínimos exigidos.

Os controlos QC" são apresentados sob a forma de uma tabela. Para cada parâmetro, estão disponíveis as seguintes informações:

- **Name** : o nome do parâmetro
- **Value** : o valor do parâmetro, incluindo :
 - Um ícone que indica se o filtro é *Passed*, *Warning* ou *Failed*
 - O valor do parâmetro para a amostra
- **Description** : Breve descrição do parâmetro

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 : Resumo das métricas de CQ relacionadas com a amostra sWGS no SomaHRD v1.2 (em cima) e no SomaHRD v1.5 (em baixo).

Os controlos de qualidade são aplicados aos parâmetros seguintes, de acordo com os limiares apresentados no final do relatório HRD (secção '**SAMPLE QUALITY CHECKS**'):

- a percentagem de células tumorais, com base apenas no valor fornecido pelo utilizador.
- cobertura média da sequenciação do genoma,
- proporção de reads corretamente alinhadas, calculada após a etapa de corte do adaptador (**apenas nas versões SomaHRD v1.2 e v1.3**),
- a métrica de confiança da instabilidade genómica (LGA e LPC) com base nos dados de cobertura (**apenas na versão SomaHRD v1.2**). Este controlo de qualidade é utilizado para medir a confiança do estado HRD devolvido. Alguns resultados positivos e negativos são marcados com uma etiqueta "low confidence" (baixa confiança) para identificar as amostras com a confiança mais baixa no estado.
- um intervalo de confiança de probabilidade HRD (*HRD probability confidence interval*)' (**somente na versão SomaHRD v1.3**). Esta métrica terá um estado *Warning* quando o erro padrão (*desvio padrão*) for superior a 0,07; terá um estado *Failed* se o 'HRD probability confidence interval' incluir o valor 0,5.
- um intervalo de confiança de probabilidade GI (*GI probability confidence interval*)' (**apenas na versão SomaHRD v1.4**). Esta métrica terá um estado *Warning* se o erro padrão (*standard deviation*) for superior a 0,07; terá um estado *Failed* se o GI probability confidence interval incluir o valor 0,5.
- uma dispersão GI (**apenas na versão SomaHRD v1.5**) que mostra a dispersão dos 100 cálculos de probabilidade GI. Esta métrica exibirá um estado de «Warning» com uma etiqueta «Moderate» ao lado do valor quando a dispersão for > 2,6 e um estado de «Failed» com uma etiqueta «High» se for > 2,7. Caso contrário, uma etiqueta «Low» aparecerá ao lado da pontuação.
- Em caso de suspeita de baixo conteúdo tumoral (*Low tumoral fraction*)' com base no sWGS, é apresentada a mensagem *WARNING* (aviso). Se não for este o caso, o controlo apresenta a mensagem *NORMAL*'.

Se qualquer uma das seguintes métricas QC estiver abaixo dos limiares de falha esperados, o resultado da análise será inconclusivo (NON CONCLUSIVE):

- '% células tumorais
- Cobertura
- % reads corretamente alinhados (**apenas para as versões SomaHRD v1.2 e v1.3**).
- dispersão GI > 2.7 **se também** $0,37 < GiP < 0,59$ (**apenas para SomaHRD v1.5**)

Neste caso, as métricas para confiança na instabilidade genómica e fração tumoral baixa não são calculadas e apresentam 'N/A' nas respectivas secções (**para a versão SomaHRD < v1.5**).

Para a versão SomaHRD 1.3, se a confiança na métrica de instabilidade genómica for *Failed*", ou seja, abaixo do limiar de falha esperado, o resultado da análise será:

- POSITIVE LOW CONFIDENCE se a probabilidade de HRD for maior ou igual a 50%.

- **NEGATIVE LOW CONFIDENCE** se a probabilidade de HRD for inferior a 50%.

Neste caso, o resultado da análise é acompanhado de um aviso.

No entanto, se a análise for acompanhada por um painel que inclua BRCA1 e BRCA2, e o painel devolver um estado BRCA positivo, o estado HRD será positivo, mesmo que uma métrica QC relativa à amostra sWGS esteja abaixo do limiar de falha (**apenas para as versões SomaHRD v1.2 e v1.3**).

Copy number variations. Esta secção é dedicada às variações do número de cópias dos genes CCNE1 e RAD51 em termos do número de cópias detetadas. Estas não são tidas em conta no cálculo da probabilidade HRD e da probabilidade GI.

Caraterísticas de instabilidade genómica. Esta secção fornece detalhes das caraterísticas que foram calculadas para determinar a instabilidade genómica do tumor. Estas caraterísticas são utilizadas para determinar a probabilidade HRD do SeqOne (**SomaHRD v1.2 e v1.3**) e a probabilidade GI (**SomaHRD versão $\geq$ v1.4**). Esta secção destina-se a fornecer uma compreensão da forma como a probabilidade é calculada. Para cada uma das caraterísticas, são fornecidas as seguintes informações:

- **Nome** : Nome da caraterística (por exemplo, LGA Status ou LGA Score)
- **Value** : O valor que foi calculado para o paciente
- **Description** : Breve descrição das caraterísticas

No caso de resultados *NÃO CONCLUSIVOS*, relacionados com métricas QC abaixo dos limiares esperados, as secções *AMPLIFICATIONS* e *GENOMIC INSTABILITY FEATURES* do relatório aparecem a cinzento.

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: Caraterísticas da probabilidade de HRD relacionadas com a instabilidade genómica.

BRCA1/BRCA2 & TP53 mutations. No caso de um teste HRD completo incluindo a análise SomaVar, fornecemos uma secção adicional dedicada às variantes patogénicas detetadas nos genes BRCA1, BRCA2 e "TP53". São listadas todas as variantes patogénicas BRCA1 e BRCA2 que correspondem aos critérios descritos na secção **Metodologia**. No entanto, apenas a variante patogénica ou provavelmente patogénica do TP53 com a frequência alélica mais

elevada será comunicada. Outras variantes TP53 patogénicas podem ser visualizadas na análise SomaVar associada. Esta secção está oculta apenas para relatórios sWGS.

As mutações detetadas são comunicadas numa tabela com as seguintes colunas:

- **Gene & Transcript:** Nome do gene HGNC e ID do transcript RefSeq
- **Variante:** Nomenclaturas HGVSC e HGVSP da variante
- **Evaluation:** Avaliação da Variant de acordo com os critérios do ACMG (*Pathogenic* ou *Likely Pathogenic*)
- **VAF:** Fração alélica da variante.

Se não tiverem sido identificadas mutações BRCA patogénicas na análise do painel associado, esta secção do relatório apresenta a seguinte mensagem: "No pathogenic mutations have been identified on BRCA1 and BRCA2 genes."

Se não tiverem sido identificadas mutações patogénicas TP53 na análise do painel associado, esta secção do relatório apresenta a seguinte mensagem: "No pathogenic mutations have been identified on TP53 gene."

BRCA1/BRCA2 & TP53 MUTATIONS

GENE & TRANSCRIPT	VARIANT	EVALUATION	VAF
BRCA1 NM_007294.4	c.3116_3117del p.Ala1039GlufsTer6	Pathogenic	60.68%
BRCA1 NM_007294.4	c.3113_3117delinsG p.Glu1038GlyfsTer9	Pathogenic	58.18%
BRCA1 NM_007294.4	c.3113_3114del p.Glu1038GlyfsTer7	Pathogenic	60.08%
TP53 NM_000546.6	c.524G>A p.Arg175His	Pathogenic	39.29%

Figura 40: Variantes patogénicas BRCA1/BRCA2 e TP53.

HRD summary. Esta última secção do relatório é dedicada aos resultados. Esta secção visa fornecer uma visão detalhada dos resultados, apresentando as seguintes informações:

- **HRD Probability' (SomaHRD v1.2 e v1.3):** O valor da probabilidade HRD, entre 0 e 1. O status é positivo quando o valor é maior ou igual a 0,5.
- **GI probability (versão SomaHRD ≥ v1.4):** O valor da probabilidade de GI, entre 0 e 1. O status é positivo quando este valor é maior ou igual a 0,5.
- **BRCA Status:** Status determinado pela deteção de pelo menos uma mutação patogénica BRCA: MUTATED, WILD-TYPE, ou "N/A" na ausência de análise do painel.
- **HRD Status ou GI Status :** POSITIVO, NEGATIVO, NEGATIVO LOW CONFIDENCE, POSITIVO LOW CONFIDENCE ou NÃO CONCLUSIVO como indicado no cabeçalho do relatório.
- **Method :** O método (ou modo) que foi utilizado para produzir a pontuação HRD:
 - HRD completa ⇒ "WGS SHALLOW + BRCA1/2 GENE PANEL"
 - Apenas superficial ⇒ "SHALLOW WGS ONLY"

No caso de resultados *NEGATIVE LOW CONFIDENCE*, *POSITIVE LOW CONFIDENCE* ou *NON CONCLUSIVE*", estes são apresentados a cinzento claro.

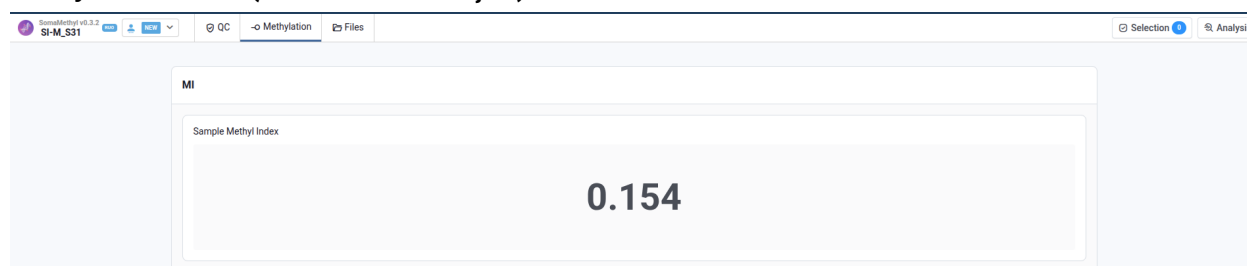
Os detalhes sobre os diferentes valores apresentados no relatório, bem como a metodologia geral do teste, estão presentes nas secções "**Footnotes**" e "**Methodology**", localizadas a seguir ao resultado.

Para além do relatório HRD, os gráficos de instabilidade genómica associados à análise HRD estão disponíveis para transferência no separador Files da análise SomaHRD. Estes gráficos representam a cobertura normalizada em todos os cromossomas dos dados sWGS.

No caso de resultados *NON CONCLUSIVE*: relacionados com métricas QC abaixo dos limiares esperados, não são gerados os QC graphs' de instabilidade genómica.

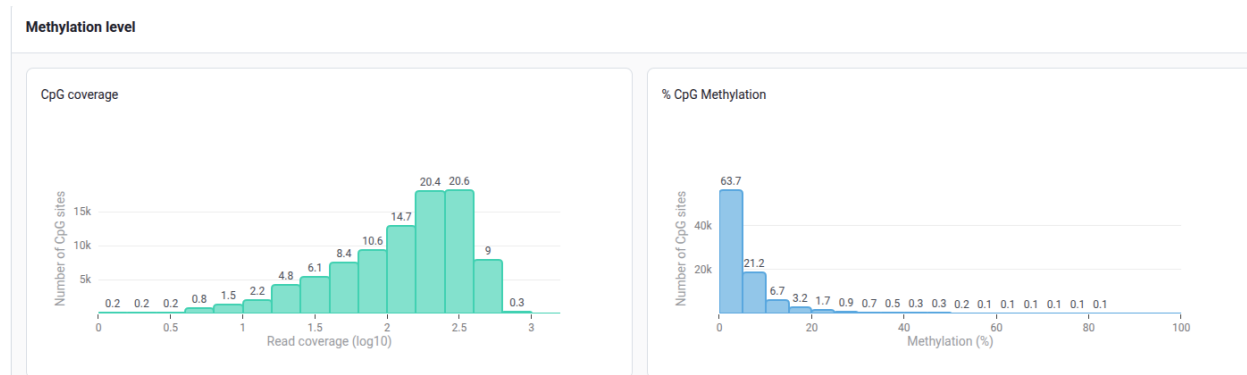
13.1.14. Dados de metilação para SomaMethyl

Methylation Index (Índice de metilação)



O Índice de metilação (MI) é uma pontuação numérica única que quantifica o estado geral de metilação de uma amostra biológica e resume o sinal de metilação associado ao tumor. Pode detectar de forma sensível até mesmo níveis baixos de DNA derivado do tumor em um fundo de DNA normal. No SomaMethyl, o MI é calculado no modo Padrão usando o algoritmo proprietário da Agilent (MethylAlgo). Quando uma coorte de referência está disponível, o MI da amostra pode ser comparado aos valores de MI da coorte e interpretado usando um limite definido pelo usuário. Para avaliar o MI, a Agilent forneceu amostras de titulação com níveis de metilação esperados. Os valores de MI calculados pelo módulo de modo Padrão do SomaMethyl são comparados aos valores de MI esperados fornecidos pela Agilent, e o desempenho é resumido usando uma métrica de concordância ponderada definida como $1 - \text{wMAPE}$ (Erro percentual absoluto médio ponderado).

Methylation level (Nível de metilação)



CpG coverage (histogram) (Cobertura de CpG (histograma))

Uma distribuição da profundidade de leitura entre os locais de CpG nas regiões-alvo (manifesto).

Eixo X (cobertura de leitura log10): cobertura por local CpG (número de leituras que suportam a chamada de metilação naquele CpG, ou seja, contagens metiladas + não metiladas).

Eixo Y (# locais CpG): número de locais CpG no manifesto que possuem essa cobertura.

Você pode verificar se os CpGs alvos estão suficientemente cobertos e se a cobertura é uniforme o suficiente para interpretar as porcentagens de metilação com confiança.

% CpG methylation (histogram) (% de metilação CpG (histograma))

Uma distribuição dos níveis de metilação entre os locais CpG nas regiões-alvo (manifesto).

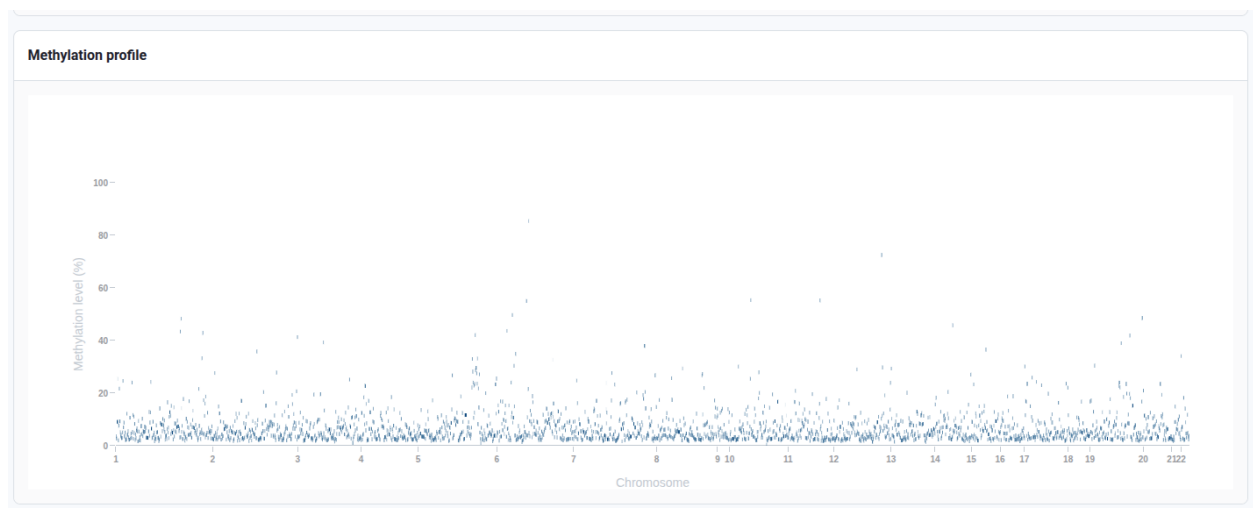
Eixo X (% de metilação): para cada local CpG, a fração de leituras metiladas entre todas as leituras que cobrem aquele CpG.

Eixo Y (# locais CpG): número de locais CpG no manifesto mostrando esse nível de metilação.

Fornece uma visão ao nível da amostra dos estados de metilação CpG em todas as regiões-alvo (frequentemente mostrando picos próximos à metilação baixa e alta, dependendo das regiões capturadas e da biologia da amostra).

Locais CpG com cobertura muito baixa produzem porcentagens de metilação menos confiáveis porque um pequeno número de leituras pode fazer o % subir ou descer. Por esse motivo, o histograma de % de metilação CpG deve sempre ser interpretado junto com o histograma de cobertura CpG (e as outras métricas de controle de qualidade).

Methylation profile (Perfil de metilação)



O gráfico do perfil de metilação mostra o sinal de metilação CpG da amostra em todo o genoma, exibido por cromossomo. Cada ponto corresponde à porcentagem média de metilação CpG computada sobre janelas genômicas sucessivas (como na saída .mcpG.track.bedGraph). Essa visualização ajuda a visualizar padrões amplos de metilação e identificar regiões com metilação maior ou menor em comparação com o resto do genoma.

Methylation Table (Tabela de metilação)

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 ▶					

A tabela de metilação lista as regiões genômicas (cromossomo:início-fim) sobre as quais o SomaMethyl resume a metilação CpG (janelas fixas ou segmentos mesclados, dependendo da etapa). Para cada região, o METH. LEVEL relata o nível de metilação agregado computado a partir das chamadas de metilação CpG nesse intervalo, e o CpG NBR. indica quantos locais CpG contribuíram para o cálculo. Cada linha é então anotada com CpG REGIONS (ilha, costa, plataforma quando aplicável) e anotações funcionais, incluindo elementos reguladores de GENOMIC FEATURE sobrepostos (promotor, intensificador, silenciador) e informações REFSEQ CDS (sequência codificadora).

14. Relatório

14.1.1. Ficheiros de saída

O separador *Files* lista os ficheiros gerados nas diferentes fases do workset. Cada ficheiro pode ser descarregado clicando no ícone  .

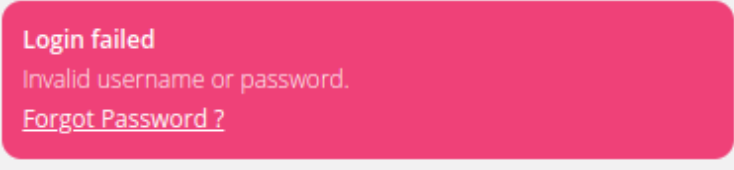
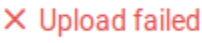
Alguns ficheiros não incluem o nome da amostra como prefixo. Para os descarregar com este prefixo, é necessário ativar a opção 'Prefix with sample name' no canto superior direito da página.

Os ficheiros estão disponíveis para descarregamento nas diferentes secções da interface:

- Os ficheiros FastQ estão disponíveis no separador *Samples* do projeto, selecionando a amostra relevante e, em seguida, o ícone  .
- Os Manifestos (ficheiros BED) estão disponíveis nas *Settings* do utilizador através do ícone no canto superior direito do painel de controlo e, em seguida, no separador *Manifests*.
- O(s) relatório(s) de análise em formato PDF está(ão) disponível(is) no separador "*Reporting*" depois de selecionar uma análise.
- Os ficheiros gerados durante uma análise (BAM, VCF, etc.) e os diferentes relatórios HTML (qualidade, técnico, etc.) estão disponíveis no separador "*Files*" para cada análise.

15. Resolução de problemas e mensagens de erro

São utilizadas várias mensagens para notificar os utilizadores de um erro

Mensagem de erro	Descrição
 A pink message box with the text: "Login failed", "Invalid username or password.", and a link "Forgot Password?".	A palavra-passe e/ou o login está/estão incorrectos. Os utilizadores são convidados a introduzir novamente o seu login e a sua palavra-passe ou a clicar em "Forgot password?" para redefinir a sua palavra-passe.
 A blue message with three dots and the text "Upload pending".	O carregamento da amostra está pendente. Os utilizadores são aconselhados a aguardar, a não fechar a janela do navegador e a não cortar a ligação à Internet.
 A red message with an 'X' icon and the text "Upload failed".	A amostra não foi carregada. <i>Este erro pode ocorrer se a janela do browser tiver sido fechada ou se a ligação à Internet tiver sofrido uma falha durante o carregamento.</i> Os utilizadores são convidados a tentar carregar novamente o(s) ficheiro(s).
 A dark blue progress bar showing 62% completion, with icons for search, document, and user.	Indica o progresso do carregamento atual Os utilizadores são aconselhados a aguardar, a não fechar a janela do navegador e a não cortar a ligação à Internet.

	Carregamento da página. Os utilizadores são convidados a aguardar.
... Bioinformatics - <i>pending</i>	A análise está pendente. Os utilizadores são convidados a aguardar.
× Failed - <i>error</i>	A análise falhou. Os utilizadores são convidados a verificar as definições e os dados utilizados para executar a análise e, em seguida, reiniciar a análise.
⚠ Something went wrong Something went wrong, you should refresh this page before continue. Feel free to contact the support if the problem persist.	Ocorreu um erro. O utilizador é convidado a clicar na tecla "Esc". Se o problema persistir, contactar o suporte em support@seqone.com.
× Missing fastq files	Esta amostra não tem os ficheiros FastQ associados, o que significa que não podem ser lançadas novas análises para essa amostra.
unexpected EOF × Verification failed	A mensagem "Unexpected EOF" (fim de ficheiro) corresponde a um ficheiro truncado. O utilizador é convidado a verificar os dados e a recarregar os ficheiros em causa.

Os utilizadores podem contactar a equipa de apoio clicando no logótipo  no canto superior direito do painel de controlo do utilizador e, em seguida, no ícone  **Support**

16. Garantia

Consulte os termos e condições da garantia no seu contrato de manutenção.

17. Aviso legal

17.1. Termos de utilização

As informações sobre a licença de utilizador podem ser acedidas clicando no logótipo da conta de utilizador seguido de "*Terms of Use*". Os utilizadores podem visualizar e descarregar as informações em formato PDF a partir da janela pop-up, clicando em "*OPEN PDF*".

17.2. Limitações de responsabilidade

A SeqOne rejeita toda e qualquer responsabilidade por qualquer utilização incorrecta ou não profissional da plataforma SeqOne.

O SeqOne é uma ferramenta de apoio à decisão clínica e não é responsável pelos diagnósticos derivados dos resultados obtidos.

Por conseguinte, os utilizadores devem ter a devida cautela ao utilizar o software e os seus resultados. Os diagnósticos devem ser confirmados através de vários exames clínicos e médicos.

17.3. Direitos de autor e marcas registadas

Todos os nomes de empresas e produtos mencionados neste guia podem ser marcas comerciais dos respectivos proprietários.