LymphoTrack® Dx Assays

Instruções de Utilização

LymphoTrack® Dx Software – MiSeq™ Versão 2.4.6

CE UKCA IVD

Este pacote de software é fornecido com cada Ensaio LymphoTrack Dx para o equipamento Illumina MiSeq com o objetivo de analisar os ficheiros FASTQ originais. O software é composto por três partes distintas: a Aplicação de Análise de Dados LymphoTrack, um plugin Repórter LymphoTrack, que é executado automaticamente após a análise e um método de Visualização LymphoTrack opcional, que consiste em quatro folhas de cálculo Microsoft Excel® XLSM. O software pode ser executado em plataformas suportadas pelo Microsoft Windows®.

Este pacote de software permite o teste de clonalidade de dados de um único alvo ou de múltiplos alvos do (Líder *IGHV* [ou "Líder"]), *IGH* FR1, *IGH* FR2, *IGH* FR3, *IGK, TRB* ou *TRG*) e, para além disso, apresenta o estado de hipermutação somática de amostras analisadas com os ensaios Líder ou *IGH* FR1.

IVD Para utilização em diagnóstico *in vitro*.

Compatibilidade do Produto

Este LymphoTrack Dx Software – MiSeq (REF 95000009) foi desenvolvido especificamente para ser utilizado e é **apenas** compatível com os seguintes ensaios da Invivoscribe:

Ref. Catálogo	Descrição	UDI
REF 91210059	LymphoTrack Dx *IGHV* Leader Somatic Hypermutation Assay Kit A - MiSeq	00850052003814
REF 91210069	LymphoTrack Dx *IGHV* Leader Somatic Hypermutation Assay Panel - MiSeq	00850052003821
REF 91210009	LymphoTrack Dx *IGH* FR1 Assay Kit A – MiSeq	00850052003647
REF 91210039	LymphoTrack Dx *IGH* FR1 Assay Panel – MiSeq	00850052003654
REF 91210089	LymphoTrack Dx *IGH* FR2 Assay Kit A – MiSeq	00850052003838
REF 91210099	LymphoTrack Dx *IGH* FR2 Assay Panel – MiSeq	00850052003845
REF 91210109	LymphoTrack Dx *IGH* FR3 Assay Kit A – MiSeq	00850052003852
REF 91210119	LymphoTrack Dx *IGH* FR3 Assay Panel – MiSeq	00850052003869
REF 91210129	LymphoTrack Dx *IGH* FR1/2/3 Assay Kit A – MiSeq	00850052003876
REF 91210139	LymphoTrack Dx *IGH* FR1/2/3 Assay Panel – MiSeq	00850052003883
REF 91220009	LymphoTrack Dx *IG*K Assay Kit A – MiSeq	00850052003906
REF 91220019	LymphoTrack Dx *IGK* Assay Panel – MiSeq	00850052003913
REF 92270019	LymphoTrack Dx *TRG* Kit A – MiSeq	00850052003678
REF 92270009	LymphoTrack Dx *TRG* Panel – MiSeq	00850052003661
REF 92250009	LymphoTrack Dx *TRB* Kit A – MiSeq	00850052003937
REF 92250019	LymphoTrack Dx *TRB* Panel – MiSeq	00850052003944

invivoscribe

Índice

1. Utilização pretendida 3
2. Glossário e abreviaturas 3
3. Requisitos mínimos do sistema 4
4. Advertências e precauções 4
5. Instalar o software 5
6. Executar o software de análise de dados 5
7. Visualização de dados utilizando o repórter LymphoTrack (automatizado, recomendado) 9
8. Visualização de dados (manual): executar as folhas de cálculo de visualização de dados de Excel (não recomendado) 10
9. Serviços técnicos e apoio ao cliente 13
10. Símbolos 13
11. Aviso legal 13
12. Perguntas frequentes sobre o software 14

1. Utilização Pretendida

O pacote LymphoTrack Dx Software – MiSeq destina-se a ser utilizado na análise de dados bioinformáticos e na visualização de ficheiros FASTQ originais gerados com o equipamento Illumina MiSeq quando usados com os Ensaios LymphoTrack Dx.

2. Glossário e abreviaturas

2.1 Glossário

Tabela 1. Termos do Glossário

Termo	Definição
Clonal	O agregado de células ou organismos geneticamente idênticos produzidos a partir de uma única célula progenitora.
Inválido	Um resultado de amostra ou de controlo que não satisfaz os critérios de validade.
Controlo de versões semântico	Um esquema de versão de software composto por 3 números (Major.Minor.Patch), em linha com o fator de risco da atualização.
Software	LymphoTrack Dx Software - MiSeq
Sistema	Todo o pacote de software, hardware e ensaio (conforme aplicável) que constituem o dispositivo médico.

2.2 Abreviaturas

Tabela 2. Definição das abreviaturas

Abreviatura	Definições
CDR3	Região determinante da complementaridade, formada pela união dos segmentos V-(D)-J e que estabelece a diversidade da resposta imunitária celular
CSV	Formato de ficheiro de valores separados por vírgulas
FASTQ	Um formato baseado em texto para armazenar uma sequência biológica (normalmente uma sequência de nucleótidos) e as suas pontuações de qualidade correspondentes
FFPE	Fixados em formalina e impregnados em parafina
IFU	Instruções de utilização
***IGH* FR1**	Cadeia pesada de imunoglobulina Quadro 1
***IGH* FR2**	Cadeia pesada de imunoglobulina Quadro 2
***IGH* FR3**	Cadeia pesada de imunoglobulina Quadro 3
IGHV	Génica da região variável da cadeia pesada de imunoglobulinas
IGK	Cadeia ligeira da imunoglobulina kappa
PCR	Reação em cadeia da polimerase
PDF	Formato de documento portátil
SHM	Hipermutação somática
TRB	Recetor de células T beta
TRG	Recetor de células T gama
TSV	Formato de ficheiro de valores separados por separadores
USB	Universal Serial Bus; permite a troca de dados e o fornecimento de energia entre vários tipos de eletrónica

3. Requisitos Mínimos do Sistema

- **Processador**: Recomenda-se Intel Core 2 Duo ou CPU mais recente.
- **Disco rígido:** São necessários pelo menos 80 GB de espaço livre no disco; recomendam-se 250 GB.
- **RAM:** Necessários 4 GB; recomendam-se 8 GB ou mais.
- **Sistema operativo:** É necessário o Windows 10 Pro ou Windows 11 Pro (64 bits).
- Um leitor de ficheiros PDF como o Adobe Acrobat Reader, para visualizar os relatórios de dados gerados pelo Repórter LymphoTrack.
- É necessário o Microsoft Office Excel 2007, 2010 ou 2013 com suporte Macro ativado para a visualização de dados.
- Uma unidade de CD-ROM.
- O LymphoTrack Dx Software – MiSeq v2.4.6 foi validado para ser utilizado com o Windows 10 Pro e Windows 11 Pro, configurado para a língua inglesa (EUA), com predefinições de visualização, e o Microsoft Excel 2007/2010/2013 para a visualização de dados alternativos. Outras versões do Windows/Excel e/ou definições de idioma/visualização podem não ser compatíveis.

4. Advertências e Precauções

- **Tamanho do ecrã predefinido**. A resolução de ecrã ótima para o LymphoTrack Dx Software – MiSeq é 100%. Aumentar o tamanho do ecrã predefinido para mais de 100% pode afetar a visualização de texto nos campos da interface do utilizador.
- **Tipo de letra do sistema**. A interface do utilizador foi desenhada de forma a utilizar as predefinições de letra do sistema num computador Windows que utilize a língua inglesa (EUA) local. Alterações nas definições de letra do sistema ou o uso de outro idioma local que não a língua inglesa (EUA) podem afetar a visualização de texto nos campos da interface do utilizador.
- **Ficheiros compatíveis.** O LymphoTrack Dx Software – MiSeq é compatível com ficheiros FASTQ.GZ gerados pelo Software Repórter MiSeq. O LymphoTrack Dx Software – MiSeq não é compatível com ficheiros FASTQ.GZ gerados pelo software *Bcl2fastq*.
- **Corte de adaptador.** A opção de corte de adaptador tem de ser selecionada ☑ ao criar folhas de amostras MiSeq, para que os ficheiros FASTQ.GZ gerados sejam cortados em função do adaptador.
- **Definições de modo suspenso ou em hibernação.** Deverá considerar desativar os modos suspenso e/ou em hibernação antes de executar a Aplicação de Análise de Dados LymphoTrack Dx. Se o computador entrar em modo suspenso ou em hibernação, o software pode deixar de funcionar.
- **Caracteres no nome do caminho e no nome do ficheiro.** 1) Evite usar espaços no nome do caminho para os ficheiros de dados ou software (os nomes de caminho incluem os nomes das pastas com os ficheiros e os nomes dos ficheiros); não é permitido mais do que um espaço consecutivo. 2) É importante que os nomes dos ficheiros contenham apenas caracteres alfanuméricos e hífen (A-Z, a-z, 0-9, - (hífen)). Se o software encontrar um carácter que não faça parte deste conjunto, pode falhar. Certifique-se de que segue esta nomenclatura para as amostras quando configurar a folha de amostras para o MiSeq. 3) Não exceda os 255 caracteres para o nome do caminho. Os nomes de amostras que fazem com que os nomes dos ficheiros de saída se aproximem deste limite são difíceis de copiar para outras localizações.
- **As amostras em duplicado devem ter um único nome.** Atribua um nome e identificador únicos a cada amostra ao designar as amostras. Se as amostras forem analisadas em duplicado, podem ser designadas por “Amostra1a” e “Amostra1b”. Se não atribuir nomes únicos às amostras a serem analisadas em conjunto na mesma lâmina de fluxo, apenas uma amostra será analisada pelo software durante o processo de análise.

5. Instalar o Software

5.1 Clique duas vezes no arquivo ***LymphoTrackDxMiSeq-2.4.6.IVD.msi*** localizado no CD do software e sigas as instruções no ecrã.

- Instale o software apenas numa unidade local e não numa unidade da rede.
- O software pode não funcionar devidamente se for executado numa ligação em rede.

6. Executar o Software de Análise de Dados

6.1 Identifique os ficheiros FASTQ.GZ cortados para adaptador e com as extremidades emparelhadas para análise. Estes ficheiros estão localizados na pasta da análise MiSeq e são os ficheiros FASTQ.GZ que serão utilizados como dados de introdução na aplicação de análise de dados **LymphoTrackDxMiSeq**.

- Os ficheiros *Undetermined_ *.fastq.gz* não são necessários para análise.

6.2 Copie os ficheiros FASTQ.GZ identificados acima para uma pasta vazia no computador. Transfira estes ficheiros para o mesmo computador local onde o Software LymphoTrack Dx está instalado.

- A unidade que contém esta pasta tem de ter no mínimo 80 GB de espaço disponível.
- Esta pasta será a **<input_folder>** (**<pasta_de_entrada>**) conforme indicado a seguir para o software.

6.3 Clique duas vezes no arquivo do aplicativo ***LymphoTrackDxMiSeq-2.4.6.IVD*** localizado na pasta **Invivoscribe** criada durante a instalação.

- O local é *Disco Local (C:) > Invivoscribe*.

6.4 Leia o acordo de licença. Para aceitar os termos, clique no botão **Accept** (**Aceitar**) para prosseguir.

6.5 Assinale o modo de análise pretendido a partir das opções nas caixas de seleção *Choose analysis mode(s) (Selecionar modo[s] de análise)*.

- O modo de análise deve corresponder ao tipo de dados na análise (Líder, *IGH* FR1, *IGH* FR2, *IGH* FR3, *IGK, TRB* e/ou *TRG*).

6.6 Selecione a notação pretendida para a marca decimal a partir das opções *Choose decimal mark (Selecionar marca decimal)*.

- Esta opção afeta posteriormente como a macro de visualização de dados do Microsoft Excel é usada e constitui uma entrada obrigatória.
- As opções são utilizar um ponto (“.”) ou uma vírgula (“,”) para a marca decimal.

IMPORTANTE! **Marca decimal.** Para determinar que opção selecionar, abra uma nova folha de cálculo no Excel e escreva **=10000** na célula A1. Clique nesta célula com o botão direito do rato e selecione **Format Cells (Formatar células)**.

- ☐ No separador *Number (Número)*, procure na lista *Category (Categoria)* e selecione **Number (Número)**. À direita, o valor para *Decimal places (Casas decimais)* é definido como **2**.
- ☐ Abaixo deste valor está uma caixa de seleção *Use 1000 Separator (Usar separador 1000)*. Assinale esta caixa e observe o número de amostras acima. Este irá conter o número original inserido, formatado de acordo com as predefinições do Excel para um número com uma marca decimal.
- ☐ Se este número aparece como *‘10,000.00’*, selecione a opção **decimal point (ponto decimal)**. Se este número aparece como *‘10.000,00’*, selecione a opção **decimal comma (vírgula decimal)**.

6.7 Clique no botão **Browse** (**Navegar**) para selecionar uma pasta de entrada. Esta pasta é a **<input_folder>** indicada acima e contém os ficheiros a serem analisados. Quando a pasta de entrada é selecionada, a mesma aparece vazia.

IMPORTANTE! A **<input_folder>** selecionada não deve conter nenhum arquivo de amostra com o mesmo nome da pasta de entrada. Se o programa foi executado com uma amostra que possuio mesmo nome de arquivo da pasta fornecida como entrada para a funcionalidade de análise, o programa irá congelar e o tempo de execução da análise aumentará significativamente (ou seja, levará 30 minutos para processar uma pequena amostra que leva segundos para processar normalmente).

6.8 Uma vez selecionada a **<input_folder>** e o modo de análise, clique no botão **Launch Program (Iniciar programa)**.

IMPORTANTE! **A pasta <input_folder> selecionada deve conter apenas ficheiros FASTQ ou FASTQ.GZ.** Se nenhum destes tipos de ficheiro nem uma mistura de tipos de ficheiros estiver presente, o programa irá apresentar uma mensagem de erro e solicitar que selecione uma pasta diferente. Se os ficheiros FASTQ forem gerados pelo *Bcl2fastq* (os nomes dos ficheiros contêm a sequência de código de barras em vez da nomenclatura "S1, S2, S3" etc.), o programa apresenta uma mensagem de erro e solicita uma nova análise dos dados utilizando o Software do Repórter MiSeq.

6.9 **Estado da análise**. Após clicar no botão **Launch Program (Iniciar programa)**, a janela principal do programa irá mostrar o estado do programa. Quando o programa terminar de analisar os ficheiros de dados de entrada, a janela principal do programa irá mostrar a mensagem *Analysis complete... (Análise concluída...)* e a localização da pasta de saída, além de solicitar que abra a folha Data Visualization (Visualização de dados) Excel apropriada.

- Uma nova análise pode ser iniciada a partir da janela principal do programa assim que a análise anterior for concluída.

IMPORTANTE! **Duração da análise.** O tamanho e o número de amostras a serem analisadas, assim como o modo de análise e o hardware do computador, irão determinar o tempo que o programa necessita para fazer a análise. Uma distribuição de leitura típica para a análise de 24 amostras é de 200 000 para um 1 milhão de leituras por amostra. Estes números podem alterar-se com as futuras versões da plataforma MiSeq.

Uma análise típica de uma amostra única de Líder *IGHV*, *IGH* FR1, *IGH* FR2, *IGH* FR3, *IGK, TRB*, ou *TRG* requer aproximadamente 1 a 2 minutos para ser concluída. Uma análise típica de 24 amostras apenas de Líder, *IGH* FR1 apenas, *IGH* FR2 apenas, *IGH* FR3 apenas, *IGK* apenas, *TRB* apenas ou *TRG* apenas requer aproximadamente 25 a 50 minutos para ser concluída. Uma análise típica de 24 amostras combinadas de Líder, *IGH* FR1, *IGH* FR2, *IGH* FR3, *IGK*, *TRB* e/ou *TRG* multiplica o tempo de análise pelo número de alvos a serem analisados.

6.10 **Pasta(s) de saída.** Dependendo do modo de análise selecionado, será criada uma ou mais pastas de saída em **<input_folder>** contendo os ficheiros FASTQ ou FASTQ.GZ de origem.

- A Tabela 3 resume a relação entre o modo de análise e o(s) nome(s) da(s) pasta(s) de saída.

Tabela 3. Nome da pasta de saída de acordo com o modo de análise

Modo de Análise	Nome(s) da(s) Pasta(s) de Saída
Líder	*LEADER_output (LÍDER_saída)*
IGH FR1	*IGH_FR1_output (IGH_FR1_saída)*
IGH FR2	*IGH_FR2_output (IGH_FR2_saída)*
IGH FR3	*IGH_FR3_output (IGH_FR3_saída)*
TRB	*TRB_output (TRB_saída)*
TRG	*TRG_output (TRG_saída)*
IGK	*IGK_output (IGK_saída)*
Qualquer combinação de: Líder, *IGH* FR1, *IGH* FR2, *IGH* FR3, *IGK*, *TRB* e *TRG*	Duas a sete das pastas de saída indicadas acima

6.11 Após o programa ter terminado a análise com sucesso, a pasta de saída irá conter vários ficheiros TSV e um ficheiro PDF para cada amostra processada.

- De notar que "*" nas tabelas a seguir representa o nome da amostra tal como introduzido na folha de amostras, seguido de informações específicas sobre a plataforma (*p.ex. samplename_S1_L001_001_combined*).
- Para as análises de Líder, *IGH* FR1, *IGH* FR2, *IGH* FR3, *TRB* e *IGK*; consulte a Tabela 4 para as descrições dos nove ficheiros fornecidos para cada amostra.

Tabela 4. Descrição dos ficheiros de saída para *IGH*, *TRB* e *IGK*.

Análise de Líder, *IGH* FR1, *IGH* FR2, *IGH* FR3, *TRB* e *IGK*	
**.fastq_read_summary.tsv* *(*.resumo_leitura_fastq.tsv)*	Contém uma lista das primeiras 200 sequências únicas, ordenadas por contagem (número de leituras), especificando: ▪ Que evento de recombinação V-J melhor se adequa ▪ Comprimento ▪ % total de leituras efetivas ▪ % total de leituras cumulativas ▪ Quão bem a leitura abrange o gene-V identificado que é alvo dos primers (%) (*IGH/IGK* e *TRB*) ▪ Putative CDR3 sequence (Sequência CDR3 putativa) Líder e *IGH* FR1 contém também: ▪ Taxa de mutação para gene-V parcial (%) Líder, *IGH* FR1, *IGH* FR2, *IGH* FR3 contêm também: ▪ Se se prevê que o rearranjo está in-frame (Y/N/n/a (S/N/n/a)). Se não é possível prever o uso de segmento do gene V ou J devido à qualidade do alinhamento, o software não será capaz de indicar se o gene está in-frame e o resultado será *n/a.* ▪ Se se prevê que o rearranjo não contém um codão stop (Y/N (S/N)). Quando não é possível fazer uma previsão de in-frame, não é possível indicar com exatidão se está presente ou não um codão stop.
**.fastq_read_summary_merged_top10_ searchtop500.tsv* *(*.resumo_leituras_fundidas_fastq_primeiras10 _procurar500primeiras.tsv)*	Mesmo formato que o ficheiro *read_summary (resumo_das_leituras)*, mas inclui apenas as 10 primeiras leituras de *read_summary (resumo_das_leituras)* fundidas com leituras da lista nas 500 primeiras sequências. Uma leitura irá fundir-se com outra se tiverem apenas 1 ou 2 nucleótidos (nts) diferentes entre si e a leitura com a contagem mais alta mantém a sua identidade de sequência, enquanto a leitura com a contagem mais baixa é adicionada à mais alta.
**.fastq_unique_reads.tsv* *(*.leituras_únicas_fastq.tsv)*	Um ficheiro FASTA de sequências únicas, ordenadas por contagem.
**.fastq_CDR3_unique_reads.tsv* *(*.leituras_únicas_fastq_CDR3.tsv)*	Um ficheiro FASTA de putative CDR3 sequences (sequências CDR3 putativas) encontradas no ficheiro FASTQ_UNIQUE_READS (LEITURAS_ÚNICAS_FASTQ).
**.fastq_VJ_sequence_frequency_family. tsv* *(*.família_frequência_ sequência_fastq_VJ. tsv)*	Cada família constitui um agrupamento de todas as recombinações possíveis encontradas numa região V ou J particular, tais como a família V_1 ou J_4. Dentro da família V_1, pode ser observado qualquer número de recombinações, tais como V_{1-69} e V_{1-8}. Este ficheiro contém uma lista de todas as recombinações da família V – família J possíveis, cada uma com uma lista ordenada de leituras (da maior à menor %). Este ficheiro é melhor utilizado num gráfico "de barras" no Excel.
**.fastq_read_summary_family.tsv* *(*.família_resumo_leituras_fastq.tsv)*	Contém uma lista das sequências únicas, ordenadas por contagem, especificando: ▪ Que evento de recombinação da família V-J melhor se adequa ▪ Comprimento ▪ % total de leituras efetivas ▪ % total de leituras cumulativas ▪ Quão bem a leitura abrange o gene-V identificado que é alvo dos primers (%) (*IGH/IGK*) ▪ Putative CDR3 sequence (Sequência CDR3 putativa) Líder e *IGH* FR1 contém também: ▪ Taxa de mutação para gene-V parcial (%)
**.fastq_read_summary_family.tsv* *(*.família_resumo_leituras_fastq.tsv)*	Líder, *IGH* FR1, *IGH* FR2, *IGH* FR3 contêm também: ▪ Se se prevê que o rearranjo está in-frame (Y/N/n/a (S/N/n/a)) ▪ Se se prevê que o rearranjo não contém um codão stop (Y/N (S/N))
**.fastq_VJ_usage.tsv* *(*.utilização_fastq_VJ.tsv)*	Uma matriz de contagens e uma matriz de % dos diferentes eventos de recombinação V-J.

Tabela 4. Descrição dos ficheiros de saída para *IGH*, *TRB* e *IGK*.

Análise de Líder, *IGH* FR1, *IGH* FR2, *IGH* FR3, *TRB* e *IGK*	
**.fastq_VJ_usage_family.tsv* *(*.família_utilização_fastq_VJ.tsv)*	Uma matriz de contagens e uma matriz de % dos diferentes eventos de recombinação da família V – família J.
**<timestamp>.pdf* *(*<carimbohora>.pdf)*	O relatório PDF gerado pelo Repórter LymphoTrack irá incluir informações semelhantes às do ficheiro *merged read summary (resumo das leituras fundidas)*, o ficheiro *read summary (resumo das leituras)*, um gráfico *V – J usage (uso de V – J)* e um gráfico de *V – J sequence frequency (frequência de sequências V – J)*. Consulte a secção 3 *Visualização de Dados Utilizando o Repórter LymphoTrack (Automatizado, Recomendado),* para obter mais informações.

Nota: *A análise de IGK* inclui ainda informações para outros elementos no locus *IGK* envolvidos em recombinações (IGKINTR e IGKDEL).
Nota: A análise de *TRB* inclui ainda informações para outros elementos no locus *TRB* envolvidos em recombinações da região D – J.
Nota: A *Putative CDR3 sequence (Sequência CDR3 putativa)* é determinada fazendo a correspondência de uma definição exata de uma sequência CDR3. Qualquer mutação em pontos de ancoragem ou falha no quadro de leitura irá resultar num valor *not found (não observado).*

- Para a análise de *TRG*, consulte a Tabela 5 para descrições dos sete ficheiros fornecidos para cada amostra.

Tabela 5. Descrição dos ficheiros de saída para *TRG*.

***Análise* de TRG**	
**.fastq_read_summary.tsv* *(*.resumo_leitura_fastq.tsv)*	Uma lista das sequências únicas, ordenadas por contagem, especificando: ▪ Que evento de recombinação V-J melhor se adequa ▪ Comprimento ▪ % total de leituras efetivas ▪ % total de leituras cumulativas ▪ Putative CDR3 sequence (Sequência CDR3 putativa)
**.fastq_read_summary_merged_top10_ searchtop500.tsv* *(*.resumo_leituras_fundidas_fastq_primeiras10_ procurar500primeiras.tsv)*	Mesmo formato que o ficheiro *read_summary (resumo_das_leituras)*, mas inclui apenas as 10 primeiras leituras de *read_summary (resumo_das_leituras)* fundidas com leituras da lista nas 500 primeiras sequências. Uma leitura irá fundir-se com outra se tiverem apenas 1 ou 2 nts diferentes entre si e a leitura com a contagem mais alta mantém a sua identidade de sequência, enquanto que a leitura com a contagem mais baixa é adicionada à mais alta.
**.fastq_unique_reads.tsv* *(*.leituras_únicas_fastq.tsv)*	Um ficheiro FASTA de sequências únicas, ordenadas por contagem.
**.fastq_CDR3_unique_reads.tsv* *(*.leituras_únicas_fastq_CDR3.tsv)*	Um ficheiro FASTA de putative CDR3 sequences (sequências CDR3 putativas) encontradas no ficheiro FASTQ_UNIQUE_READS (LEITURAS_ÚNICAS_FASTQ).
**.fastq_VJ_sequence_frequency.tsv*	Uma lista de todas as recombinações da V-J possíveis, cada uma com uma lista ordenada de leituras (da maior à menor %). O primeiro número representa a % total de leituras que a primeira leitura única contabilizou em todas as leituras sequenciadas. Este ficheiro é melhor utilizado num gráfico "de barras" no Excel.
**.fastq_VJ_usage.tsv*	Uma matriz de contagens e uma matriz de % dos diferentes eventos de recombinação V-J.
**<timestamp>.pdf*	O relatório PDF gerado pelo Repórter LymphoTrack irá incluir informações semelhantes às do ficheiro *merged read summary (resumo das leituras fundidas),* o ficheiro *read summary (resumo das leituras),* um gráfico *VJ usage (uso de VJ)* e um gráfico de *VJ sequence frequency (frequência de sequências VJ).* Consulte a secção 3 *Visualização de Dados Utilizando o Repórter LymphoTrack (Automatizado, Recomendado)*.

Nota: A *Putative CDR3 sequence (Sequência CDR3 putativa)* é determinada fazendo a correspondência de uma definição exata de uma sequência CDR3. Qualquer mutação em pontos de ancoragem ou falha no quadro leitura irá resultar num valor *not found* (*não observado*).

- A(s) pasta(s) de saída contêm relatórios PDF e ficheiros de dados de saída TSV.
- Os ficheiros TSV podem ser utilizados como entradas na(s) folha(s) de cálculo de visualização de dados.

7. Visualização de Dados Utilizando o Repórter LymphoTrack (Automatizado, Recomendado)

7.1 É criado automaticamente um relatório PDF para cada amostra após a análise ser concluída.

- Este relatório terá um nome derivado do nome da amostra presente na folha de amostras, seguido de informações específicas sobre a plataforma, e terminando com o registo da data e hora (DDMMAAhhmmss).

IMPORTANTE! **Atenção ao gravar o relatório PDF:** Pode ser alterada uma ou mais colunas dependendo do programa utilizado para abrir o ficheiro PDF. NÃO altere nenhuma secção do ficheiro PDF Se o utilizador suspeitar que foram feitas alterações, NÃO grave o ficheiro PDF. Se o utilizador suspeitar que foram gravadas alterações, repita a análise a fim de criar um novo relatório PDF.

7.2 A primeira página do relatório PDF irá conter um valor de ***Index Q30 (Q30 do índice)***, calculado como a percentagem de bases com um score de qualidade igual ou superior a 30, mas apenas de leituras atribuídas a um índice após o corte. Este valor pode ser utilizado para determinar se o índice foi aprovado na especificação padrão do sistema Illumina.

- Em alguns casos, o score Q30 da análise fornecido pelo sequenciador irá subestimar a qualidade de uma análise devido à contagem indeterminada de sequências e regiões não cortadas na direção do Q30.
- As regiões indeterminadas e não cortadas são normalmente de qualidade inferior e não contribuem para a subsequente análise pelo LymphoTrack.

7.3 A primeira página do relatório PDF irá incluir uma tabela com os resultados do ***merged read summary (resumo das leituras fundidas)***, seguida dos gráficos com os resultados da ***V – J Sequence Frequency (Frequência de sequências V – J)*** e ***V – J Usage (Uso de V – J)***. As páginas remanescentes contêm uma tabela com os resultados do ***read summary (resumo das leituras)*** não fundidas.

- O conteúdo destas será detalhado na Tabela 6:

Tabela 6. Descrição do conteúdo do relatório PDF.

Ficheiro PDF de Saída para o Repórter LymphoTrack	

Read Summary
(Resumo das leituras)
	Contém uma lista das primeiras 200 sequências únicas, ordenadas por contagem, especificando:
	• Que evento de recombinação V – J melhor se adequa a cada sequência
	  ○ Sequência
	  ○ Comprimento das sequências
	  ○ % total de leituras efetivas
	• % total de leituras cumulativas listadas para as 200 primeiras leituras
	  ○ Quão bem a leitura abrange o gene-V identificado que é alvo dos primers (%) (*IGH/IGK*)
	  ○ Putative CDR3 sequence (Sequência CDR3 putativa)
	
	Líder e *IGH* FR1 contém também:
	• Taxa de mutação para gene-V parcial (%)
	
	Líder, *IGH* FR1, *IGH* FR2, *IGH* FR3 contêm também:
	• Se se prevê que o rearranjo está in-frame (Y/N/n/a (S/N/n/a)). Se não é possível prever o uso de segmento do gene V ou J devido à qualidade do alinhamento, o software não será capaz de indicar se o gene está in-frame e o resultado será *n/a.*
	• Se se prevê que o rearranjo não contém um codão stop (Y/N (S/N)). Quando não é possível fazer uma previsão de in-frame, não é possível indicar com exatidão se está presente ou não um codão stop.

Merged Read Summary *(Resumo das leituras fundidas)*	Mesmo formato que a folha de cálculo *Read Summary (Resumo das leituras)*, mas apenas para as primeiras 10 leituras após a fusão com as 500 primeiras leituras se as leituras tiverem apenas 1 ou 2 nts diferentes entre si. Adicionalmente, o score Q30 do índice é apresentado na *Total Read Count (Contagem total de leituras).*
V – J Sequence Frequency Graph *(Gráfico de frequência de sequências V – J)*	Um histograma que mostra as 200 primeiras frequências das sequências para cada recombinação de família V – família J (para *IGH/IGK* e *TRB*) ou gene V – gene J (para *TRG*).

Tabela 6. Descrição do conteúdo do relatório PDF.

Ficheiro PDF de Saída para o Repórter LymphoTrack	
V – J Usage Graph *(Gráfico de uso V – J)*	Um gráfico de barras 3D que mostra o uso da família V – família J para as 200 primeiras sequências (para *IGH*/*IGK* e *TRB*) ou o uso do gene V – gene J para todas as sequências (para *TRG*).

Nota: A *Putative CDR3 sequence (Sequência CDR3 putativa)* é determinada fazendo a correspondência de uma definição exata de uma sequência CDR3. Qualquer mutação em pontos de ancoragem ou falha no quadro leitura irá resultar num valor *not found* (*não observado*).

Nota: A análise de *IGK* inclui ainda informações para outros elementos no locus *IGK* envolvidos em recombinações (IGKINTR e IGKDEL).

Nota: A análise de *TRB* inclui ainda informações para outros elementos no locus *TRB* envolvidos em recombinações da região D – J.

8. Visualização de dados (manual): Executar as Folhas de Cálculo de Visualização de dados de Excel (não recomendado)

8.1 Verifique as definições de Segurança no Microsoft Excel a fim de verificar se o suporte Macro está devidamente ativado:

8.1.1. Excel 2007: Clique no botão **Office** e selecione **Excel Options (Opções do Excel)**.
8.1.2. Excel 2010/2013: Clique no menu **File (Ficheiro)** e selecione **Options (Opções)**.
8.1.3. Na janela *Options (Opções)*, clique em **Trust Center** (**Centro de Confiança**) e, em seguida, no botão **Trust Center Settings...** (**Definições do Centro de Confiança...**).
8.1.4. Clique em **Macro Settings (Definições Macros)** e defina Disable all macros with notification (Desativar todas as macros com notificação).
8.1.5. Clique em **OK** e novamente em **OK**.
8.1.6. Feche o Excel.

IMPORTANTE! Como os arquivos de visualização podem ser salvos como uma pasta de trabalho (o que removeria a capacidade de funcionamento da macro), copie os arquivos de visualização para um caminho de arquivo diferente antes de iniciar as análises.

8.2. Clique duas vezes no ficheiro da folha de cálculo de *Visualização de Dados* adequado para o abrir.

- Os ficheiros de visualização estão localizados na subpasta Visualizations (Visualizações) localizada na pasta de instalação do software. Estão disponíveis os seguintes ficheiros de folha de cálculo de visualização:
  - ☐ *LymphoTrackDxIGHVisualization.xlsm* para Líder, *IGH* FR1, *IGH* FR2, *IGH* FR3
  - ☐ *LymphoTrackDxIGKVisualization.xlsm* para *IGK*
  - ☐ *LymphoTrackDxTRBVisualization.xlsm* para *TRB*
  - ☐ *LymphoTrackDxTRGVisualization.xlsm* para *TRG*

8.3. Ao abrir o ficheiro, pode aparecer a mensagem: ***Security Warning*** - ***Macros have been disabled*** (***Alerta de Segurança - As macros foram desativadas***).

8.3.1. Excel 2007: Clique no botão **Options... (Opções...)** à direita do alerta e selecione **Enable this content (Ativar este conteúdo)** e clique em **OK**.
8.3.2. Excel 2010/2013: Clique no botão **Enable Content (Ativar Conteúdo)**.

- A macro de visualização de dados irá iniciar automaticamente.

8.4. Será mostrado o acordo de licença, que deverá ser aceite para utilizar o software.

- Depois de aceitar o acordo de licença, abre-se uma caixa de diálogo com uma mensagem: Please select the Read Summary file for the sample you want. (Selecione o ficheiro do resumo das leituras para a amostra que pretende.)
- A macro só permite a visualização de dados de uma amostra de cada vez, pelo que terá de repetir este procedimento para cada amostra que pretende visualizar.

8.5. Navegue até à pasta de saída criada pela aplicação de análise de dados que contém os dados a serem visualizados e selecione o ficheiro ***FASTQ_UNIQUE_READS*** (***LEITURAS_ÚNICAS_FASTQ***) para a amostra a analisar.

8.6. Clique no botão **Open** (**Abrir**).

- A macro irá transferir os ficheiros de dados para a amostra selecionada e criar várias folhas de cálculo, cada uma com diferentes visualizações legíveis dos dados.
- Isto pode demorar algum tempo, dependendo do tamanho da amostra e do hardware do computador.

IMPORTANTE! Devido a uma limitação do Microsoft Excel, o software de visualização não é capaz de mostrar mais do que 1 048 576 linhas de dados por folha. Serão mostradas apenas as primeiras 1 048 576 linhas de dados se qualquer um dos ficheiros de saída da análise de dados para a amostra atual exceder este limite de linhas.

Uma vez atingido este limite, o programa irá mostrar uma mensagem que indica a(s) folha(s) afetada(s) por esta limitação. Clique em **OK** para prosseguir.

Embora nem todos os dados possam ser mostrados, os gráficos não são afetados e estarão corretos.

IMPORTANTE! Os gráficos são gerados assumindo um sentido de visualização **Left-to-right (Da esquerda para a direita)**. Em países onde o sentido de leitura normal é "da direita para a esquerda", as predefinições do Excel podem ser incompatíveis e os gráficos não estarão corretos. Para resolver este problema, a opção de visualização do Excel deve ser alterada para **Left-to-right (Da esquerda para a direita)**. Esta opção pode ser definida em *Office > Excel Options (Office > Opções do Excel)* no Excel 2007 ou em *File > Options (Ficheiro > Opções)* no Excel 2010/2013. Selecione **Advanced (Avançadas)** e percorra para baixo até às definições de *Display (Visualização)*. Mude a opção *Default direction (Direção predefinida)* para **Left-to-right (Da esquerda para a direita)** e, em seguida, clique em **OK**. Da próxima vez que o Excel for iniciado, as folhas de cálculo serão lidas da esquerda para a direita e o software de visualização irá gerar gráficos corretamente.

Caso tal suceda e se for necessária assistência adicional, contacte o departamento de Assistência Técnica da Invivoscribe através do número +1 (858) 224-6600, de segunda a sexta, 7:00 – 17:00 (Verão: GMT - 9 horas/Inverno: GMT - 8 horas), ou por e-mail para support@invivoscribe.com.

8.6.1. Se aparecer uma mensagem relativa à *Complex Formatting (Formatação complexa)*, clique em **Yes (Sim)**.

8.7. **Gravar o ficheiro de visualização**. Quando o processamento for concluído, irá aparecer uma mensagem a informar que o programa foi concluído. Recomenda-se vivamente que grave o ficheiro neste momento.

IMPORTANTE! Ao gravar o ficheiro, certifique-se de que seleciona **Save As (Guardar como)** e selecione um Livro do Excel (XLXS ou XLS), porque o procedimento por defeito ao selecionar **Save (Guardar)** ou clicar no botão de disco será substituir a Macro XLSM pelo livro atual. Isto irá requerer que a macro seja reinstalada com uma nova cópia do CD do software para análises subsequentes.

8.7.1. Se aparecer uma mensagem de aviso sobre gravar o ficheiro como um livro sem macros, clique no botão **Yes (Sim)**.

8.8. O livro completo para Líder, *IGH* FR1, *IGH* FR2, *IGH* FR3, *IGK, TRB* e *TRG* incluirá as seguintes oito folhas de cálculo descritas na Tabela 7.

Tabela 7. Folhas de cálculo incluídas no ficheiro Excel de saída.

Livro Completo para Líder, *IGH* FR1, *IGH* FR2, *IGH* FR3, *IGK, TRB* e *TRG*	
Read Summary (Resumo das leituras)	Contém uma lista das primeiras 200 sequências únicas, ordenadas por contagem, especificando: ▪ Que evento de recombinação V-J melhor se adequa a cada sequência ▪ Sequência ▪ Comprimento das sequências ▪ % total de leituras efetivas ▪ % total de leituras cumulativas para todas as leituras (*TRG* apenas) ▪ % total de leituras cumulativas listadas para as 200 primeiras leituras (*IGH*/*IGK* e *TRB*) ▪ Quão bem a leitura abrange o gene-V identificado que é alvo dos primers (%) (*IGH*/*IGK*) ▪ Putative CDR3 sequence (Sequência CDR3 putativa)

Tabela 7. Folhas de cálculo incluídas no ficheiro Excel de saída.

Livro Completo para Líder, *IGH* FR1, *IGH* FR2, *IGH* FR3, *IGK, TRB* e *TRG*	
Read Summary *(Resumo das leituras)*	Líder e *IGH* FR1 contém também: ▪ Taxa de mutação para gene-V parcial (%) Líder, *IGH* FR1, *IGH* FR2, *IGH* FR3 contêm também: ▪ Se se prevê que o rearranjo está in-frame (Y/N/n/a (S/N/n/a)). Se não é possível prever o uso de segmento do gene V ou J devido à qualidade do alinhamento, o software não será capaz de indicar se o gene está in-frame e o resultado será *n/a.* Se se prevê que o rearranjo não contém um codão stop (Y/N (S/N)). Quando não é possível fazer uma previsão de in-frame, não é possível indicar com exatidão se está presente ou não um codão stop. ▪ Putative CDR3 sequence (Sequência CDR3 putativa).
Merged Read Summary *(Resumo das leituras fundidas)*	Mesmo formato que a folha de cálculo *Read Summary (Resumo das leituras),* mas apenas para as primeiras 10 leituras após a fusão com as 500 primeiras leituras se as leituras tiverem apenas 1 ou 2 nts diferentes entre si.
VJ Sequence Frequency Graph *(Gráfico de frequência de sequências VJ)*	Um histograma que mostra as 200 primeiras frequências das sequências para cada recombinação de família V – família J (para *IGH*/*IGK* e *TRB*) ou gene V – gene J (para *TRG*).
VJ Usage *(Uso de VJ)*	Contagens originais e frequências das 200 primeiras recombinações de família V - família J (para *IGH*/*IGK* e *TRB*) ou gene V – gene J (para *TRG*).
VJ Usage Percent Graph *(Gráfico de percentagem de uso de VJ)*	Um gráfico a 3D que mostra as 200 primeiras frequências das sequências para cada recombinação de família V – família J (para *IGH/IGK* e *TRB*) ou gene V – gene J (para *TRG*).
VJ Usage Raw Graph *(Gráfico original de uso de VJ)*	Um gráfico a 3D que mostra as 200 primeiras contagens originais para cada recombinação de família V – família J (para *IGH/IGK* e *TRB*) ou gene V – gene J (para *TRG*).
VJ Sequence Frequency *(Frequência de sequências VJ)*	Uma lista "em pilha" de todas as possíveis recombinações da família V – família J (para *IGH*/*IGK* e *TRB*) ou gene V – gene J (para *TRG*), cada uma com uma lista ordenada das primeiras 200 leituras (da maior à menor %).
Unique Reads *(Leituras únicas)*	Um ficheiro FASTA de sequências únicas, ordenadas por contagem.

Nota: *IGH* listado na tabela abaixo refere-se à análise de Líder, *IGH* FR1, *IGH* FR2 e *IGH* FR3. Os gráficos de frequências para estes alvos génicos apenas especificam *IGH* no título. A pasta de dados de saída especifica se o Líder ou uma das regiões da framework (FR) foi analisada pelo software e esta informação é incluída no campo de localização do ficheiro de dados no topo da folha de cálculo Read Summary (Resumo das leituras).

Nota: A análise de *IGK* inclui ainda informações para outros elementos no locus *IGK* envolvidos em recombinações (IGKINTR e IGKDEL).

Nota: A análise de *TRB* inclui ainda informações para outros elementos no locus *TRB* envolvidos em recombinações da região D – J

Nota: Primeiras 200: Quando especificado acima, as folhas de cálculo irão mostrar as primeiras sequências (até 200) para Líder, *IGH* FR1, *IGH* FR2, *IGH* FR3, *TRB* e *IGK*. Todas as sequências são mostradas para *TRG* em Read Summary (Resumo das leituras).

Nota: A *Putative CDR3 sequence (Sequência CDR3 putativa)* é determinada fazendo a correspondência de uma definição exata de uma sequência CDR3. Qualquer mutação em pontos de ancoragem ou falha no quadro de leitura irá resultar num valor *not found (não observado).*

IMPORTANTE! As atribuições de genes V/J encontradas no Read Summary (Resumo das leituras) são, por vezes, marcadas como *none* (*nenhum*). Estas atribuições baseiam-se na qualidade do melhor alinhamento entre a leitura e os genes de referência. Se o melhor alinhamento está abaixo de um limite mínimo de qualidade, a atribuição do gene não pode ser feita com confiança e é marcada como *none* (*nenhum*).

IMPORTANTE! A taxa de mutação para o gene-V parcial (%) irá mostrar 'n/a' para IGH FR2 e FR3. O alcance do gene-V é limitado devido aos tamanhos de amplicons mais reduzidos para estes alvos, o que resulta em cálculos inexatos da taxa de mutação.

- ☐ Ao analisar o estado de hipermutação somática das amostras, deverá focar-se nas 4 primeiras leituras (limite as amostras monoclonais ou biclonais às primeiras 2 ou 4 leituras, respetivamente), visto que as análises de taxa de hipermutação somática (SHM) e de in-frame/codão stop são as mais informativas para estas leituras. Devido à natureza imparcial das plataformas de NGS, são sequenciadas e analisadas muitas mais leituras do que as 4 primeiras, a maioria das quais são rearranjos pouco abundantes, ou podem ser atribuídas a artefactos da amplificação e sequenciação. As análises de SHM e in-frame/codão stop não são úteis em leituras não clonais nestes casos. De notar que a Invivoscribe utiliza o GRCh37 para sequências de referência e os resultados podem diferir ligeiramente comparado com a base de dados IMGT. Além disso, muitos pseudogenes e genes não funcionais não são incluídos na análise. Deve interpretar com cuidado rearranjos com um estado de hipermutação somática igual ou próximo do limite de 2%, em particular quando realizar o ensaio de *IGH* FR1, que é apenas direcionado para uma porção da região IGHV, visto que a sequência a montante do local de ligação do primer não será avaliada.

9. Serviços Técnicos e Apoio ao Cliente

Damos valor ao seu negócio. Temos todo o gosto em auxiliá-lo/a na compreensão deste software e em fornecer-lhe assistência técnica contínua, de segunda a sexta-feira, para manter o desempenho eficiente dos ensaios no seu laboratório.

Informações para Contacto

Invivoscribe, Inc
10222 Barnes Canyon Road | Building 1 | San Diego | California 92121-2711 | EUA
Telefone: +1 858 224-6600 | Fax: +1 858 224-6601 | Horário de funcionamento: 9:00 – 17:00 GMT - 8/GMT - 7
Serviços Técnicos: support@invivoscribe.com | Apoio ao Cliente: sales@invivoscribe.com | Website: www.invivoscribe.com

10. Símbolos

Os seguintes símbolos são utilizados na rotulagem dos produtos de NGS da Invivoscribe.

REF	Referência do catálogo		Data de validade
VOL	Volume de reagente	EC REP	Representante autorizado na União Europeia
LOT	Número de lote		Consulte as instruções de utilização
	Condições de armazenamento	IVD	Para utilização em diagnóstico *in vitro*
UDI	Identificador de Dispositivo Exclusivo		Fabricante
UK CA	Conformidade do Reino Unido Avaliada	UKRP	Pessoa responsável no Reino Unido
CH REP	Representante Autorizado Suíço	CE	Conformidade Europeia
	Aplicação de software		

11. Aviso Legal

Para aceder aos Avisos Legais relacionados com este produto, visite: https://invivoscribe.com/legal-notices/

12. Perguntas Frequentes Sobre o Software

P1: **Por que motivo o Read Summary (Resumo das leituras) da análise de *IGH*, *IGK* ou *TRB* tem menos de 200 leituras?**

R1: O LymphoTrack Dx Software - MiSeq alinha as 200 primeiras leituras durante a análise, mas o *Read Summary (Resumo das leituras)* apenas inclui as leituras que foram alinhadas com um gene V ou um gene J encontrado na referência. Se uma leitura não puder ser alinhada com a referência, esta é excluída de Read Summary (Resumo das leituras).

Os motivos por trás de um não alinhamento com a referência são vários, mas normalmente concentram-se num problema com a amostra, resultando na ocorrência de uma grande quantidade de amplificação não específica. Este evento pode derivar de números reduzidos de células-alvo na amostra, condições de PCR inadequadas ou tipos de amostras de baixa qualidade tais como tecidos fixados em formalina e impregnados em parafina fragmentados. A fim de reduzir a probabilidade de ocorrência destes eventos, siga precisamente as Instruções de Utilização.

P2: **Por que motivo a atribuição-V ou a atribuição-J indica *none (nenhum)* na tabela com o Read Summary (Resumo das leituras)?**

R2: A fim de se poder atribuir uma leitura a um gene V e/ou gene J específico, deve atingir-se um limite de qualidade mínimo durante o alinhamento. Se o alinhamento é fraco, não pode ser feita uma atribuição com confiança e tal é marcado **none** (nenhum). De notar que uma leitura não pode ser marcada **none** (**nenhum**) para ambos os genes V e J, visto que estes não seriam comunicados, de acordo com a condição discutida na pergunta 2.

P3: **O *V–J Sequence Frequency Graph (Gráfico de frequência de sequências V–J)* tem valores insólitos para o eixo y. A *% total reads (% total de leituras)* é da ordem dos milhões. Como é que isto é possível?**

R3: Tal deve-se provavelmente ao facto de executar o programa LymphoTrack Dx utilizando o tipo de marca decimal errada para a sua região. Consulte a secção 6 *Executar o Software de Análise de Dados*, especificamente para a seleção da marca decimal.

P4: **Todas as folhas de cálculo dos gráficos estão vazias, mas as folhas de cálculo sem gráficos têm dados. Para onde foram os gráficos?**

R4: Isto pode dever-se ao facto de a direção predefinida do programa Excel ser *right-to-left (da direita para a esquerda)* em vez de *left-to-right (da esquerda para a direita)*. Se os títulos das colunas numa folha de cálculo vazia começam em **A** no lado direito da folha, terá de mudar a direção para *left-to-right (da esquerda para a direita)* antes de usar a macro de visualização de dados. Para alterar esta definição, escolha *Excel options (Opções do Excel)* no botão **Office** no canto superior esquerdo, clique em **Advanced (Avançadas)** e desça até à secção *Display (Visualização)*. Aqui irá encontrar uma opção denominada *Default direction (Direção predefinida)* e ao selecionar **Left-to-right (Da esquerda para a direita)** poderá visualizar os dados das análises corretamente. Nota: Isto altera de forma permanente a direção até que seja novamente alterada. Encerrar o Excel não reverte esta alteração.

P5: **O programa bloqueou depois de clicar em *'Launch Program' ('Iniciar programa')* e apresentou este erro:**

Application terminated with error: Filename not compatible. Please use only MiSeq MSR software generated FASTQ files. (Aplicação terminou com erro: o nome do ficheiro não é compatível. Utilize apenas os ficheiros FASTQ gerados pelo software MiSeq MSR.)

For help with this error, please contact Invivoscribe Technical Support. (Para assistência com este erro, contacte a Assistência Técnica da Invivoscribe.)

R5: O LymphoTrack Dx software – MiSeq prevê ficheiros FASTQ com extremidades emparelhadas gerados pelo software Repórter MiSeq (MSR) incluído na máquina MiSeq. Se o software MSR não tiver sido utilizado para gerar ficheiros de sequências, a chamada de base tem de ser realizada utilizando o software MSR. Consulte o Guia do Utilizador do MiSeq para obter mais informações.

Realizar a chamada de base utilizando o software MSR é importante porque, embora outros programas possam gerar resultados muito semelhantes aos do software MSR (por exemplo, Bcl2fastq), estes programas também podem gerar resultados muito diferentes, dependendo dos parâmetros específicos. O Ensaio LymphoTrack Dx só é válido aquando da análise de ficheiros FASTQ gerados a partir do software MSR.

P6: **Por que motivo os resultados da análise parecem incluir menos leituras do que o esperado?**

R6: Se existem poucas ou nenhumas leituras totais analisadas de acordo com o Read Summary (Resumo das leituras), verifique se foi selecionado o tipo de análise correto (Líder, *IGH* FR1, *IGH* FR2, *IGH* FR3, *IGK, TRB* ou *TRG*) quando executar o LymphoTrack Dx software – MiSeq. Se parece haver um número reduzido de leituras totais, compare o tamanho dos ficheiros FASTQ a serem analisados com o tamanho dos ficheiros FASTQ presentes no equipamento MiSeq. Se o equipamento MiSeq tiver ficheiros maiores, então os dados de sequenciação devem ser transferidos novamente a fim de garantir a análise de todos os conjuntos de dados.

O sequenciador MiSeq mostra mensagens diferentes quando processa os dados de sequenciação. Uma mensagem pode indicar *Sequencing Complete (Sequenciação concluída)* e a outra depois indicar *Analysis Complete (Análise concluída)*. A transferência de dados FASTQ entre estas duas mensagens pode resultar numa análise LymphoTrack Dx incompleta.

P7: **Por que motivo o Read Summary (Resumo das leituras) contém uma leitura na lista das dez primeiras que não está no Merged Read Summary (Resumo das leituras fundidas)?**

R7: A leitura "desaparecida" pode ter sido fundida com uma leitura classificada acima da mesma, ou é possível que o alinhamento desta leitura tenha sido excecionalmente fraco. Durante o processo de fusão, as leituras no resumo das leituras são alinhadas com as do ficheiro de leituras únicas. No entanto, as leituras no resumo das leituras foram orientadas para cumprirem o alinhamento com a referência e, possivelmente, têm a orientação anti-sentido comparativamente ao gene de referência. Isto acontece apenas quando o alinhamento é muito fraco e quando um alinhamento do mesmo sentido ou anti-sentido é quase equivalente, sendo o alinhamento anti-sentido ligeiramente melhor. Em casos como estes, é melhor ignorar completamente a leitura.

IMPORTANTE: **Utilize apenas leituras do *Merged Read Summary (Resumo das leituras fundidas)* para determinar evidência de clonalidade.**

P8: **Por que motivo a análise demora tanto tempo (mais do que uma hora por amostra)?**

R8: O ficheiro de registos apresenta o tempo que está a demorar cada passo. Após a linha ***INFO: Decompressing files*** *('INFO: Descompactar ficheiros')*, existe uma outra linha que indica quantos ficheiros e quanto tempo demorou a descompactação. Este número está em milissegundos. Se a descompactação está a demorar várias horas, verifique se os ficheiros de destino estão num servidor de ficheiros. Tipicamente, qualquer processo I/O intensivo, como a descompactação, será mais lento numa rede. Se possível, copie os ficheiros para o computador local, ou faça a descompactação dos ficheiros diretamente antes de executar o software. Certifique-se de que os ficheiros compactados e descompactados não estão na mesma pasta. O software não irá prosseguir se ambos os ficheiros FASTQ e FASTQ.GZ estiverem na pasta de destino.

Se os passos subsequentes de parsing e análise estiverem a demorar muito tempo, os ficheiros FASTQ poderão ser muito grandes. Exemplos de ficheiros FASTQ grandes incluem ficheiros obtidos quando se incluem menos de 24 índices numa única análise com o MiSeq e, como resultado, as amostras sequenciadas têm um maior alcance. Um segundo exemplo seria se a biblioteca estiver desequilibrada e incluir uma maior proporção de uma amostra com um índice versus a proporção de amostras com outros índices. O tempo de análise previsto nestas Instruções de Utilização assumem uma análise de 24 índices com bibliotecas equilibradas.

P9: **O Q30 da análise reportado pelo equipamento MiSeq não cumpre a especificação do sistema Illumina. Se a análise LymphoTrack for executada e os controlos aparentarem ter tido o desempenho previsto, será que a análise falhou?**

R9: O Q30 da análise é calculado em tempo real durante o processo de chamada de base. Neste momento, não foi executado qualquer outro processamento de dados, como demultiplexação e corte de adaptador. Isto significa que qualquer sequência de um índice de qualidade fraco ou que ocorra fora da sequência de adaptadores está a contribuir para o score Q30. A sequência detetada fora da sequência de adaptadores é normalmente de baixa qualidade e este cenário ocorre frequentemente com dímeros de primers. Se uma biblioteca tiver um nível suficiente de dímeros de primers, tal pode provocar uma redução no Q30 da análise. Como verificação secundária, os resultados do Repórter LymphoTrack apresentam o score Q30 do índice conforme calculado a partir de todas as leituras cortadas nesse índice. Se um score Q30 do índice também não cumprir a especificação do sistema Illumina, considere que esse índice falhou.

P10: **Porque é que as 10 principais leituras combinadas são diferentes quando os mesmos dados são reanalisados?**

R10: As 10 principais contagens de leituras combinadas podem variar ligeiramente quando os dados são analisados várias vezes devido à ordenação das sequências quando existem várias sequências com exatamente a mesma contagem de leituras. Nos casos em que várias sequências têm contagens de leitura idênticas, o software pode ordená-las e classificá-las aleatoriamente. As diferenças observadas devem ser insignificantes e não afetar a interpretação dos resultados.