

SEQUENCIACÃO DE DNA – NGS

MANUAL DE BOAS PRÁTICAS

FERNANDO REGATEIRO
COORDENAÇÃO

SOFIA MARQUES
SONYA NETO
MIGUEL PINHEIRO
MANUEL SANTOS
GABRIELA MOURA

LABORATÓRIO DE MEDICINA DO GENOMA
IBIMED, UNIVERSIDADE DE AVEIRO



IMPRESA DA
UNIVERSIDADE
DE COIMBRA
COIMBRA UNIVERSITY PRESS

1 | 5

© NOVEMBRO 2023. Imprensa da Universidade de Coimbra.

Autor: Sofia Marques, Sonya Neto, Miguel Pinheiro, Manuel Santos, Gabriela Moura

Título: Sequenciação De Dna – Ngs: Manual de Boas Práticas

Responsável Científico: Fernando Regateiro

Edição

Imprensa da Universidade de Coimbra

Email: imprensa@uc.pt

URL: http://www.uc.pt/imprensa_uc

Vendas online: <http://livrariadaimprensa.uc.pt>

Coordenação editorial

Maria João Padez de Castro

Design: Carlos Costa

Infografia: Imprensa da Universidade de Coimbra

Execução gráfica: KDP

ISBN: 978-989-26-2524-9

eISBN: 978-989-26-2525-6

DOI: <https://doi.org/10.14195/978-989-26-2525-6>

Cofinanciado por:

CENTRO 

 PORTUGAL
2020

 UNIÃO EUROPEIA
Fundos Europeus
Estruturais e de Investimento

Centro 08-5864-FSE-000039

FERNANDO REGATEIRO

COORDENAÇÃO

SOFIA MARQUES

SONYA NETO

MIGUEL PINHEIRO

MANUEL SANTOS

GABRIELA MOURA

SEQUENCIAÇÃO DE DNA - NGS

MANUAL DE BOAS PRÁTICAS

Imprensa da Universidade de Coimbra

Coimbra University Press

(Página deixada propositadamente em branco)

SEQUENCIACÃO DE DNA – NGS

MANUAL DE BOAS PRÁTICAS

FERNANDO REGATEIRO
COORDENAÇÃO

SOFIA MARQUES
SONYA NETO
MIGUEL PINHEIRO
MANUEL SANTOS
GABRIELA MOURA

LABORATÓRIO DE MEDICINA DO GENOMA
IBIMED, UNIVERSIDADE DE AVEIRO

I|U **IMPRESA** DA
UNIVERSIDADE
DE COIMBRA
COIMBRA UNIVERSITY PRESS

(Página deixada propositadamente em branco)

Sumário

| | |
|---|-----------|
| Introdução | 9 |
| | |
| Capítulo 1 – Processamento laboratorial..... | 11 |
| 1. Bioética e consentimento..... | 13 |
| 1.1. Comitês de ética..... | 13 |
| 1.2. Dados genéticos..... | 14 |
| 2. Procedimentos operativos | 16 |
| 3. Colheita, transporte e receção de amostras biológicas..... | 16 |
| 4. Extração de gDNA..... | 18 |
| 4.1. Manipulação de amostras biológicas | 18 |
| 4.1.1. Avaliação de risco..... | 19 |
| 4.1.2. Grupos de risco..... | 21 |
| 4.1.3. Níveis de segurança biológica..... | 21 |
| 4.1.4. Normas para manipulação de amostras clínicas..... | 22 |
| 4.2. Controlo de qualidade gDNA..... | 24 |
| 4.3. Armazenamento gDNA..... | 26 |
| 5. Preparação de bibliotecas..... | 27 |
| 5.1. Metodologias..... | 28 |
| 5.2. Etapas do fluxo de trabalho | 29 |
| 5.3. Controlo de qualidade das bibliotecas..... | 31 |
| 5.4. Desafios na preparação de bibliotecas | 32 |

| | |
|---|-----------|
| 6. Sequenciação | 36 |
| 6.1. Etapas da sequenciação | 36 |
| 6.2. Monitorização da corrida..... | 37 |
| 6.3. Desafios na sequenciação | 38 |
| 7. Automatização de processos | 39 |
| | |
| Capítulo 2 – Análise e gestão de dados | 41 |
| 1. Requisitos técnicos: software & hardware | 41 |
| 2. Processamento de dados | 42 |
| 2.1 Visão geral do pipeline..... | 43 |
| 2.2 Pré-processamento | 44 |
| 2.3 Alinhamento | 46 |
| 2.4 Identificação de variantes de nucleotídeo único, inserção e deleção no DNA | 47 |
| 2.5 Identificação de variação no número de cópias no DNA..... | 47 |
| 2.6 Anotação e filtragem de variantes | 47 |
| 2.7 Interpretação e classificação de variantes..... | 48 |
| 3. Armazenamento de dados..... | 49 |
| | |
| Capítulo 3 – Qualidade..... | 53 |
| 1. Sistemas de gestão da qualidade | 53 |
| 2. Critérios de aceitação e rejeição | 55 |
| 3. Programas de controlo de qualidade | 56 |
| 3.1. Controlo de qualidade interno | 56 |
| 3.2. Controlo de qualidade externo | 56 |
| 4. Benchmarking para identificação de variantes..... | 58 |
| | |
| Bibliografia | 61 |

Introdução

A sequenciação de nova geração (NGS) permite a sequenciação paralela de um elevado número de fragmentos de ácido desoxirribonucleico (DNA) e tem possibilitado avanços rápidos em diferentes áreas da investigação fundamental e na prática clínica. Esta novidade científica e tecnológica foi também acompanhada pelo aumento da complexidade laboratorial e por uma maior exigência na análise bioinformática dos dados gerados. Apesar dos custos de sequenciação por amostra decrescerem à medida que a tecnologia é aprimorada e são lançados novos equipamentos, o NGS continua a ser considerado uma tecnologia dispendiosa e não acessível a todos os laboratórios, sobretudo no contexto de investigação.

O sucesso da sequenciação é determinado pela correta preparação das amostras e análise dos dados gerados. Qualquer erro no processo traduz-se na baixa qualidade e fiabilidade dos resultados, e consequentemente à perda de tempo e dinheiro, pelo que o avanço e capacitação técnica para a sequenciação genómica deve incluir a formação adequada dos elementos envolvidos e a adoção de boas práticas técnicas e metodológicas.

Este guião de boas práticas não pretende impor uma plataforma, kit ou protocolo de sequenciação em particular. O presente documento tem antes como objetivo final contribuir para a produção de dados genómicos e clínicos de qualidade, nos laboratórios da Região Centro.

Foi realizado com um foco particular no processamento e análise de variantes germinativas. Todavia, outras omicas têm lugar no âmbito da medicina genómica que, apesar das diferentes finalidades e proveniências, são, de algum modo, também abrangidas pelas recomendações contidas neste documento. O mesmo se aplica às diferentes metodologias. Apesar de atualmente existirem no mercado várias plataformas de NGS, será evidenciada a tecnologia de sequenciação por síntese (SBS), utilizada pelas plataformas de sequenciação Illumina, que é também a mais difundida na Região Centro. Contudo, cabe a cada laboratório a escolha dos métodos mais ajustados à sua rotina laboratorial.

Este manual está dividido em três capítulos distintos: (i) Processamento Laboratorial; (ii) Análise e Gestão dos Dados e (iii) Gestão da Qualidade.

PROCESSAMENTO LABORATORIAL

O processo laboratorial compreende várias etapas, necessárias para a preparação das amostras para sequenciação (Figura 1). Algumas integram um fluxo de procedimentos transversais a todos os tipos de amostras primárias e inclusive, a todos os protocolos que possam ser utilizados. Os próximos subcapítulos serão focados nas etapas standard da sequenciação genómica de amostras humanas para a sequenciação dirigida de painéis de genes, exomas (WES) e genomas completos (WGS).

A utilização de protocolos comerciais, previamente testados e validados, é a opção mais segura e simplificada, mesmo num contexto de investigação. Contudo, no que respeita à prática clínica, as opções são mais restritas. Os testes de diagnóstico devem seguir a marcação CE-IVD (diretiva CE 98/79/EC). Esta diretiva corresponde a um esquema regulatório para equipamentos médicos de diagnóstico *in vitro* (IVD) destinados ao mercado europeu, estando em vigor como um requisito legal obrigatório desde 2003. Genericamente, aplica-se a equipamentos, reagentes, instrumentos e softwares utilizados na análise de amostras humanas cujo intuito é a obtenção de dados relativos ao estado fisiológico ou patológico de indivíduos. O principal objetivo é a garantia de qualidade, segurança e reprodutibilidades dos testes.

Ainda na prática clínica, estão disponíveis na literatura algumas valiosas recomendações publicadas por alguns colégios e sociedades de Genética, como o Colégio Americano de Genética Médica (ACMG) [1] e a Sociedade Europeia de Genética Humana (ESHG) [2]. Os laboratórios que produzem testes de diagnóstico devem trabalhar na direção destas recomendações. Sugere-se também a consulta de outras recomendações adicionais, como as da Administração de Alimentos e Medicamentos (FDA) [3]. Estas recomendações aplicam-se tanto ao processamento laboratorial, como à análise e reporte dos dados obtidos e são ainda uma ferramenta importante na implementação e validação do NGS nos laboratórios clínicos. Num contexto de investigação, embora díspar da prática clínica, os laboratórios devem tender também para estas mesmas recomendações, com o objetivo de produzirem dados de elevada qualidade e de grande reprodutibilidade. Podem ainda ser consultadas recomendações publicadas por várias sociedades e grupos de trabalho de diferentes países, como o Colégio Canadano de Médicos Geneticistas (CCMG) [4] ou a Sociedade Holandesa para o Diagnóstico Laboratorial em Genética Clínica (VKGL) [5]. Contudo, será de notar que algumas das recomendações foram criadas no contexto particular dos sistemas de saúde de cada país, podendo não se aplicar diretamente ao contexto nacional.

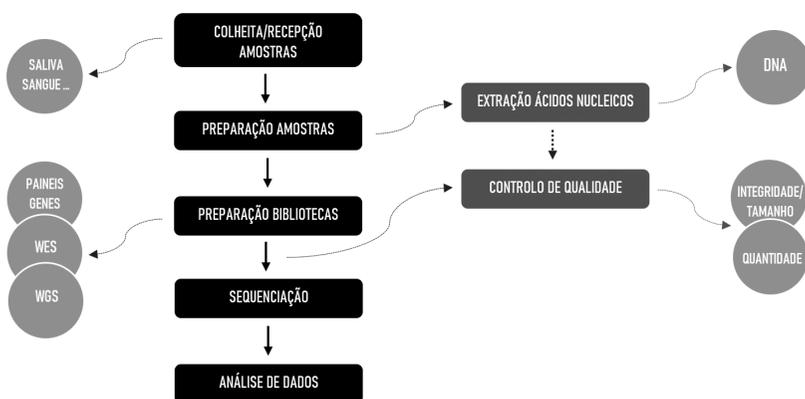


Figura 1 – Fluxograma da sequenciação de amostras de DNA por NGS, desde a colheita e/ou recepção das amostras primárias até à obtenção dos dados para posterior análise bioinformática.

1. Bioética e consentimento

A declaração Universal sobre Bioética e Direitos Humanos, emitida pela UNESCO [6] em 2005, apelou a que os Estados-membros, perante os avanços científicos e tecnológicos, desenvolvessem esforços para incorporem, nos procedimentos associados a esses avanços, os princípios do respeito pelos direitos humanos e liberdades fundamentais, enunciados nesta declaração.

Entre os princípios apresentados, o consentimento (Artigo 6º e 7º), prévio, livre e esclarecido da pessoa em causa é obrigatório para que possa ser realizado um estudo científico e/ou intervenção médica, incluindo com carácter de diagnóstico. Também as informações obtidas não devem ser utilizadas ou difundidas para outros fins que não aqueles para que foram coligidos ou consentidos (Artigo 9º). A declaração determina ainda que a pessoa em causa pode retirar o seu consentimento a qualquer momento e por qualquer razão, sem que daí resulte qualquer desvantagem ou prejuízo para a mesma. Caso o estudo seja realizado em pessoas incapazes de exprimir o seu consentimento, a investigação só deve ser realizada quando se antevê um benefício direto para a saúde da pessoa em causa ou, excecionalmente, se efetuada no interesse da saúde de outras pessoas pertencentes à mesma categoria. Contudo, deve ser compatível com a proteção dos direitos individuais da pessoa e deve ser respeitada a recusa em participar na investigação. De acordo com o Artigo 3º, os interesses e o bem-estar do indivíduo devem prevalecer sobre o interesse exclusivo da ciência ou da sociedade.

1.1. Comitês de ética

Os comités de ética são órgãos coletivos, multidisciplinares e pluralistas, peças-chave no contexto da Bioética, sendo responsáveis pela avaliação das questões éticas relevantes na presente área., desde a esfera pessoal da ética à esfera pública/política. Estes grupos são imprescindíveis para apoiar as instituições e profissionais nesta vertente, tendo também um papel ativo na divulgação da bioética ao público em geral. Na esfera política, auxiliam na formulação de novas recomendações e na regulação normativa.

No âmbito internacional, existem vários conselhos de ética, como o Grupo Europeu de Ética em Ciência e Novas Tecnologias (GEE), designado pela Comissão Europeia, que auxilia nos aspetos éticos das políticas Europeias (Parlamento Europeu e Conselho Europeu), e a Comissão de Bioética do Conselho da Europa (DH-BIO).

Porém, todos os Estados-membros da União Europeia (EU) têm um Conselho Nacional de Ética ou uma estrutura equivalente. Em Portugal, o Conselho Nacional de Ética para as Ciências da Vida (CNECV) é, desde 1990 (Lei n.º 14/90, de 9 de junho) o órgão de garantia nacional da bioética [7]. Este órgão independente, funciona junto da Presidência do Conselho de Ministros, e tem como competência, segundo a mesma lei, a análise sistemática, e emissão de pareceres, dos problemas morais suscitados pelos progressos científicos nos domínios da biologia, da medicina ou da saúde em geral. Adicionalmente, este órgão é responsável pela apresentação de um relatório anual ao Primeiro-Ministro sobre o estado da aplicação das novas tecnologias à Vida Humana e respetivas implicações de natureza ética e social, formulando as recomendações que tenha por convenientes [7].

1.2. Dados genéticos

O rápido avanço da Genética, e a acelerada produção de dados neste domínio em particular, tornou necessária a criação de orientações práticas de ação, que permitam dominar melhor as mudanças observadas e previstas. Desta forma, em 2004, foi aprovada, pela Organização das Nações Unidas para a Educação, Ciência e Cultura (UNESCO), a Declaração Internacional sobre Dados Genéticos Humanos [8], que veio prolongar a Declaração Universal sobre o Genoma Humano e os Direitos Humanos de 1997 [9]. Esta declaração teve como objetivo a garantia do respeito pela dignidade humana e a proteção dos direitos humanos e das liberdades fundamentais, em matéria de recolha, tratamento, utilização e conservação de dados genéticos humanos [8].

Ao longo dos anos subsequentes, foram emitidos protocolos e recomendações adicionais. Em 2008 e anos subsequentes, foi revista a Declaração de Helsínquia da Associação Médica Mundial de forma a enunciar os princípios éticos para a investigação médica em seres

humanos, incluindo o uso de material humano e dados. Em 2016, foi adotado o Regulamento Geral sobre a Proteção de Dados (RGPD) da EU, que estabelece as regras relativas à proteção das pessoas singulares no que diz respeito ao tratamento de dados pessoais e à livre circulação desses dados (Regulamento UE 2016/679) [10], com execução assegurada em Portugal pela Lei n.º 58/2019, de 8 de agosto [11]. No contexto de investigação em saúde, de realçar a iniciativa lançada em 2015 pela Infraestrutura de Investigação, Biobancos e Recursos Biomoleculares (BBMRI-ERIC), e reiniciada em 2020, com vista à elaboração de um código de conduta para a investigação em saúde, com o objetivo de simplificar a interpretação do RGPD e a sua aplicação neste contexto.

Com os rápidos avanços científicos e tecnológicos, e tendo em conta as perspetivas futuras, urge a necessidade de reflexão e revisão constante de recomendações e princípios normativos. Entre as temáticas-alvo, está a partilha de dados genéticos. Em 2020, é divulgada uma comunicação da Comissão Europeia ao Parlamento Europeu, com uma estratégia europeia para a gestão de dados, incluindo dados relativos à saúde. Nesse documento, a Comissão apoiou o estabelecimento de espaços comuns europeus de dados, incluindo um espaço de dados de saúde, que dê suporte ao progresso científico e aos sistemas de saúde. Pretende ainda desenvolver medidas legislativas e não legislativas setoriais para o espaço europeu de dados de saúde e implementar as necessárias infraestruturas de dados, os instrumentos e a capacidade computacional [12].

Em Portugal, algumas leis e decretos foram publicados no contexto de dados genéticos e da Bioética. A Lei n.º 12/2005 de 26 de janeiro [13], que define o conceito de informação de saúde e genética, a circulação de informação e a intervenção sobre o genoma humano no sistema de saúde, bem como as regras para a colheita e conservação de produtos biológicos para efeitos de testes genéticos ou de investigação [13]. A Lei n.º 21/2014 de 16 de abril [14], relativa à Investigação Clínica, e o Decreto-Lei n.º 80/2018 de 15 de outubro [15], que estabelece os princípios e regras aplicáveis às comissões de ética que funcionam nas instituições de saúde, nas instituições de ensino superior e em centros de investigação biomédica que desenvolvam investigação clínica.

2. Procedimentos operativos

As etapas que constituem o processo laboratorial devem ser uniformizadas e devidamente detalhadas, com recurso a procedimentos operativos (PO) descritivos e acessíveis a todos os colaboradores envolvidos no processo. Estes procedimentos não devem ser estanques. Devem ser revistos periodicamente e adaptados à evolução de cada laboratório. As alterações, tal como a sua criação, devem ser registadas, datadas e validadas, antes de serem difundidas pelos colaboradores.

Os PO são elementos fundamentais nos sistemas de qualidade dos laboratórios, que serão abordados no último capítulo deste guião.

Todas as etapas descritas nos tópicos seguintes devem ser alvo dos PO, desde a colheita/receção de amostras até à análise e gestão dos dados, detalhada no Capítulo II.

3. Colheita, transporte e receção de amostras biológicas

Os profissionais que realizam as colheitas, armazenamento e transporte de amostras biológicas humanas devem respeitar todas as regras de segurança de forma a garantir a sua segurança e a dos restantes colaboradores, bem como garantir a proteção da comunidade e do ambiente. Sempre que é detetada uma não-conformidade, esta deve ser registada e comunicada.

A colheita deve ser realizada por um profissional com formação adequada para o efeito, com recurso a material esterilizado e apropriado à natureza da amostra. Todas as amostras devem estar devidamente etiquetadas, com os dados do paciente e a data e hora da colheita. Os formulários com o pedido do teste ou outras informações devem acompanhar as amostras, todavia, devem permanecer à parte em envelopes ou sacos, preferencialmente impermeáveis. Se houver registo ou transferência de dados pessoais deve ser respeitado o RGPD.

De acordo com o Despacho n.º 10009/2019 de 5 de novembro (Manual de Boas Práticas Laboratoriais de Patologia Clínica ou Análises Clínicas) [16], as amostras devem ser conservadas e transportadas em condições apropriadas às suas características e aos exames a realizar, definidas pelo diretor técnico do laboratório/diretor de serviço. O transporte deve ser efetuado em cumprimento do regulamento nacional e internacional específico para o transporte de produtos biológicos, que para todos os efeitos devem ser considerados potencialmente infecciosos. O transporte deve ainda ser realizado de forma a minimizar a possibilidade de queda, colisão ou derramamento, seja no transporte realizado entre locais ou já dentro do próprio laboratório.

As amostras devem ser transportadas com recurso a recipientes selados e estanques, preferencialmente de plástico e com tampas de rosca. Se possível, devem ser utilizadas bandejas de contenção. Caso o transporte ocorra entre diferentes edifícios devem ser utilizados recipientes secundários de transporte rígidos, como caixas, que devem ser regularmente descontaminadas e devidamente etiquetadas, incluindo o símbolo de risco biológico.

Para efeitos de transporte, a substância potencialmente infecciosa a transportar deve ser subclassificada nas diferentes categorias com base na patogenicidade dos microrganismos que contém, ou pode conter, devendo igualmente ser atribuído o número de identificação apropriado, regulado pela Organização das Nações Unidas (ONU). As amostras humanas podem ser classificadas em categoria A, categoria B ou isentas, sendo da categoria A as substâncias que apresentam maiores riscos de biossegurança e, por isso, sujeitas a medidas mais restritas de transporte. Nas três categorias deve ser utilizado um sistema de embalagem tripla que inclui o recetáculo primário impermeável e estanque (por exemplo, tubo de recolha de sangue) envolto em material absorvente para prevenção de um potencial vazamento. O recetáculo primário deve ser introduzido numa embalagem secundária também impermeável e, por fim, protegidos por uma terceira camada, uma embalagem protetora adequada às condições de transporte que, sendo necessário, pode conter também o material refrigerante como, por exemplo, gelo seco.

Para mais detalhes, sugere-se a consulta das recomendações da ONU sobre o transporte de mercadorias perigosas, que inclui a lista das classificações usadas no transporte destas amostras [17] e a orientação da Organização Mundial de Saúde (OMS) para o transporte de substâncias [18]. No caso das amostras da categoria A, devem ser também seguidos os requisitos adicionais ao sistema de embalagem tripla da ONU que estão descritos na instrução P620, e para a categoria B, os requisitos descritos na instrução P650 [17].

4. Extração de gDNA

Os protocolos para o isolamento de DNA genómico (gDNA) devem ser adaptados à espécie utilizada e à sua aplicação subsequente. Por exemplo, o isolamento de DNA de bactérias não segue o mesmo protocolo utilizado para isolar DNA humano, nem o nível de integridade do DNA que será necessário para um protocolo de sequenciação de fragmentos curtos é igual ao de fragmentos longos. Contudo, independentemente do protocolo utilizado, o objetivo é claro: a obtenção de um gDNA puro e com qualidade, mas sem comprometimento da segurança. Antes da manipulação das amostras biológicas, o operador deve ter bem presente todas as regras de biossegurança gerais e específicas das amostras, para sua proteção individual, dos técnicos e investigadores, mas também do ambiente e da comunidade em geral. A proteção alcançada por estas normas reduz a probabilidade de exposição, mas também minimiza o impacto da exposição acidental não evitada, por garantir uma adequada contenção da exposição.

4.1. Manipulação de amostras biológicas

As recomendações e regras de biossegurança presentes neste capítulo não dispensam a consulta dos manuais de biossegurança disponibilizados por cada instituição parceira. Recomendamos também a consulta do Manual de Biossegurança laboratorial da OMS [19] e as recomendações particularizadas desta mesma organização para alguns organismos, metodologias e técnicas, conforme disponibilizadas na respetiva

página web [20]. A leitura deste, e doutros documentos equivalentes, é particularmente importante para as instituições que planeiam a criação de novos espaços, a fim de conseguirem criar um espaço o mais adaptado possível às suas necessidades atuais e idealmente preparado para rapidamente cumprir necessidades futuras que incluam medidas de contenção reforçadas.

As instituições devem assegurar a formação inicial e continua dos colaboradores sobre as normas e regras de segurança básica dos procedimentos e do próprio manual de segurança de cada instituição, e os mesmos devem ser revistos com frequência. A formação inicial pode ser feita com recurso a cursos e formações internas e a formação continua, por exemplo, através da realização periódica de seminários e workshops. O mesmo se aplica às regras de eliminação de resíduos, que devem ser realizadas conforme a legislação em vigor, e às regras particulares de segurança em caso de acidente com agentes infecciosos, como em caso de derrame ou exposição. As instituições devem ainda adotar estratégias que assegurem que todos os colaboradores estão informados sobre as regras e que leram de facto os PO e manuais, realizando, por exemplo, provas/testes individualizados.

Os contactos de emergência e os equipamentos de segurança e de uso em caso de acidente, como é o caso do kit de primeiros socorros, kit de derrame, extintores, estação de lavagem de olhos, chuveiros, entre outros, devem estar devidamente identificados, em locais apropriados para o efeito e acessíveis a todos os colaboradores. Todos os incidentes e acidentes devem ser notificados aos responsáveis institucionais e alvo de investigação apropriada, que deve culminar na implementação de ações corretivas.

Apesar de existirem várias causas associadas à exposição acidental a agentes biológicos patogénicos humanos, este capítulo foca-se no uso apropriado de equipamentos de proteção individual (EPI), na correta manipulação das amostras biológicas e nos requisitos físicos essenciais.

4.1.1. Avaliação de risco

A manipulação das amostras biológicas deve ser realizada de acordo com sua classe de biossegurança (classe de risco 1 ao 4). Contudo a

OMS, na 4ª edição do Manual de Biossegurança laboratorial (2021) [19], ressalva que o grupo de risco não corresponde diretamente ao nível de biossegurança de um laboratório. O risco real é influenciado, não só pelo agente que está a ser manipulado, como também pelo procedimento, pela competência técnica e pelas instalações/recursos do laboratório. Dado que o risco reflete a probabilidade e consequências de ocorrer um incidente, sugerem a realização de uma avaliação de risco baseada em evidências com a implementação de medidas de segurança ajustadas à realidade individual dos laboratórios.

Infelizmente, os laboratórios nem sempre são construídos ou equipados para situações particulares, mas sim para fazer face a um conjunto de possibilidades procedimentais, alteráveis ao longo do tempo, particularmente no caso dos laboratórios em contexto de investigação. Adicionalmente, a falta de recursos financeiros e humanos qualificados e de infraestruturas adequadas em alguns institutos pode levar a dificuldades na gestão do esquema de avaliação de risco.

Na avaliação de risco, os laboratórios devem reunir informação sobre os agentes biológicos com que pretendem trabalhar, os procedimentos e equipamentos que utilizarão, e os recursos técnicos e tecnológicos que estão disponíveis. Devem questionar como pode ocorrer um incidente (por exemplo, exposição ao agente), a sua probabilidade de ocorrer e as consequências se tal acontecer. Devem ainda perceber quais seriam os riscos aceitáveis e os não aceitáveis, e se os últimos podem ser controlados. Após estas etapas, deve ser desenvolvida uma estratégia de controlo de risco, que pode passar por perceber se os recursos disponibilizados são suficientes para alcançar e manter as medidas de controlo de risco necessárias. Após seleção e implementação, as medidas devem ser revistas periodicamente, tal como o próprio risco. Para mais detalhes recomenda-se a consulta da monografia de avaliação de risco associada à 4ª edição do Manual de Biossegurança laboratorial (2021) [19]. Para mais informações sobre regras de proteção dos trabalhadores contra os riscos de exposição a agentes biológicos durante o trabalho deve ser consultado o Decreto-Lei n.º 84/97, de 16 de abril [21], revisto e alterado pelo Decreto-Lei n.º 102-A/2020, de 9

de dezembro [22], que estabelece as prescrições mínimas de proteção da segurança e da saúde.

4.1.2. Grupos de risco

Apesar das atuais recomendações da OMS, que vão no sentido de um maior distanciamento entre grupos de risco e níveis de segurança biológica dos laboratórios, esta correspondência continua enraizada na rotina laboratorial, pelo que, nas próximas secções, será na mesma abordado este tema tendo em conta a dicotomia grupos de risco vs níveis de segurança biológica.

Tendo em conta esta classificação, o Grupo de Risco 1 corresponde ao manuseamento de um microrganismo sem risco ou de baixo risco individual e coletivo. O Grupo de Risco 2 corresponde a agentes patogénicos de risco individual moderado e risco coletivo baixo. A exposição a estes agentes pode causar doença no operador, mas existem tratamentos eficazes e medidas que previnem o risco de propagação para a comunidade e o ambiente. No caso do Grupo de Risco 3, está-se na presença de um agente patogénico de risco individual alto e risco coletivo baixo. O Grupo de Risco 4 corresponde a um alto risco individual e coletivo. A exposição a estes agentes causa geralmente doença grave e nem sempre possui tratamento eficaz e medidas que previnam o risco de propagação, podendo transmitir-se facilmente para a comunidade [23].

4.1.3. Níveis de segurança biológica

Os laboratórios de segurança biológica podem designar-se por laboratórios de Nível 1, 2, 3 e 4, tendo por base um conjunto de características e normas para o manuseamento de agentes biológicos de vários grupos de risco, mas não descurando outros fatores que influenciam o nível de segurança biológica. O Nível 1 e 2 corresponde ao nível base de segurança biológica. O Nível 3 será um de laboratório de confinamento e o Nível 4 de confinamento máximo. Para esclarecimentos sobre as características das instalações de qualquer nível de segurança biológica deve ser consultada a 3ª edição do Manual de Segurança Biológica em laboratório (2004) [13].

4.1.4. Normas para manipulação de amostras clínicas

No caso de amostras clínicas, para as quais geralmente não há informação adequada, estas devem ser manuseadas, no mínimo, de acordo com as normas e procedimentos para um Nível de segurança biológica 2. Em geral, estas instalações permitem o manuseamento de agentes pertencentes ao Grupo de Risco 2, contudo, em determinadas circunstâncias, após uma avaliação de risco, poderá ser necessário um nível de segurança mais elevado, com maior confinamento de aerossóis. O laboratório de segurança biológica 2 fornece as condições para serviços básicos de saúde, diagnóstico e pesquisa. Este laboratório deve ser de uso exclusivo do pessoal autorizado, permanecendo sempre com as portas fechadas e sinalizadas com o símbolo de risco biológico.

A manipulação de amostras biológicas deve respeitar todas as boas práticas de procedimentos microbiológicos e deve ser realizada dentro de uma câmara de segurança biológica (CSB), alvo de certificação, manutenção e re-certificação regular. As CBS devem ser ligadas, pelo menos 5 minutos antes do início da utilização, e permanecer ligadas 5 minutos após o termo da mesma, de forma a garantir que o ar contaminado é expelido do interior. A superfície de trabalho e paredes interiores devem ser descontaminadas antes e depois de cada utilização, assim como todo o material e equipamento reutilizável. O cloro e os álcoois são dos agentes descontaminantes mais utilizados nos laboratórios de genética, contudo, o agente manipulado pode influenciar a exigência da descontaminação, pelo que a escolha do produto utilizado (antisséptico, germicida químico ou desinfetante), deve ser feita com atenção. Caso a CSB possua lâmpadas ultravioletas (UV), estas devem ser ligadas durante 20 minutos no fim da utilização, para descontaminação.

Durante a utilização, os movimentos de introdução e retirada dos braços na CSB devem ser cuidadosos e minimizados, de forma a não interferirem com o fluxo de ar, e deve ser evitada a passagem de outros colaboradores por trás do operador. As grelhas frontais de entrada e traseira destas câmaras não podem ser bloqueadas com papel (nenhum documento deve ser colocado no interior), equipamentos ou outros objetos, e os equipamentos geradores de aerossóis devem ser dispostos no

fundo da câmara. Deve ainda evitar-se o uso de chamas vivas. Para a centrifugação de amostras infecciosas fora da CSB, devem ser utilizados copos herméticos de segurança centrífuga, enchidos e esvaziados dentro da CSB. Preferencialmente, devem ser utilizados tubos e frascos com tampa de rosca e material descartável.

Em caso de derrame de amostras biológicas, a superfície afetada deve ser imediatamente limpa com a câmara a funcionar. O derrame deve ser coberto com papel absorvente e desinfetante e limpo após alguns minutos, da periferia para o centro, repetindo as vezes que for necessário, dependendo da sua extensão. Caso ocorra um derramamento de alto risco, devido à grande quantidade de aerossóis gerados ou grande volume de líquido derramado, o laboratório deve ser abandonado por um período que permita a dissipação dos aerossóis e que as partículas mais pesadas assentem. Os laboratórios devem ter todo o procedimento para utilização das CSB detalhado num PO, incluindo o procedimento necessário em caso de derrame.

As CSB são divididas em três classes diferentes, sendo que cada uma fornece diferentes níveis de proteção. A classe II é ainda diferenciada em subclasses (A1, A2, B1 e B2). Apesar de todas potenciarem a proteção do utilizador, nem todas garantem a proteção do material manuseado ou possibilitam o manuseamento de radionucleotídeos ou químicos voláteis. A CSB de classe I fornece proteção pessoal e pode igualmente ser utilizada para trabalhar com radionucleotídeos ou químicos voláteis. Contudo, não assegura a proteção do material manuseado, uma vez que o ar injetado para a superfície de trabalho é aspirado diretamente da sala. A CSB de classe II fornece proteção pessoal e ao material, pois o ar injetado na superfície de trabalho é esterilizado por filtro de HEPA. Esta câmara permite a manipulação de agentes pertencentes aos Grupos de Risco 2 e 3 ou Risco 4, quando utilizados com fatos de pressão positiva. As diferentes subclasses da classe II possuem diferenças ao nível dos sistemas de injeção do ar (velocidade) e pressão, bem como do sistema de recirculação e exaustão do ar, sendo que a expulsão pode ser realizada para a sala ou para o exterior. A expulsão para o exterior permite a manipulação de radionucleotídeos ou químicos voláteis. A CSB da classe III fornece o nível de proteção mais elevado, sendo utilizado para a manipulação de agentes

pertencentes ao Grupo de Risco 4, e apropriadas para os laboratórios de Níveis 3 e 4 de segurança biológica. O ar é expelido através de um sistema distinto depois de passar por dois filtros HEPA e o acesso à superfície de trabalho é feito com luvas de borracha conectadas à entrada da câmara. Para a manipulação de amostras biológicas podem ser utilizadas CSB de classe I ou II, contudo, a opção de Classe II com sistema de exaustão distinto torna-se a opção mais polivalente para os laboratórios.

Todo o material, antes de ser reutilizado ou descartado, deve ser descontaminado, por esterilização em autoclave ou incineração (no caso de descarte), devendo existir um autoclave, ou outros meios de descontaminação, próximos do laboratório. Mesmo após a descontaminação, os resíduos não podem ser colocados em aterros. No caso de reutilização, a limpeza do material só pode ser feita após a descontaminação.

Os colaboradores devem estar equipados com bata ou fato de proteção e luvas apropriadas a cobrir os punhos da bata e, sempre que necessário, com outros dispositivos de proteção adicionais como aventais plásticos, óculos, viseiras, manguitos, protetores de sapatos ou máscaras. O equipamento de proteção individual não pode ser utilizado fora do laboratório. É ainda recomendado que os colaboradores que manipulem amostras biológicas estejam vacinados contra a Hepatite B, já que um dos principais riscos de manipulação de amostras biológicas humanas é o risco de infeção por este agente e por HIV.

As pipetas e pontas utilizadas para a manipulação de amostras biológicas devem ter filtro, de forma a evitarem a contaminação desse equipamento.

Para mais informações sobre os princípios de biossegurança que devem ser implementadas, incluindo as regras de desinfeção e esterilização, deve consultar-se a 3ª edição do Manual de Segurança Biológica em laboratório (2004) [23].

4.2. Controlo de qualidade gDNA

A qualidade do gDNA isolado é fundamental para o sucesso da sequenciação e determinação do sucesso da extração. É nesta fase que o técnico/ investigador confirma se o DNA obtido está dentro dos parâmetros aceites para a fase seguinte – preparação das bibliotecas para sequenciação.

Devem, por isso, ser definidos pelo laboratório os critérios de aceitação e rejeição de uma amostra de DNA. Estes critérios acabam, a maior parte das vezes, por ser definidos pelo kit/protocolo utilizado na preparação das bibliotecas. Se o laboratório, por experiência, alterar estes critérios, as alterações devem ficar registadas num PO acessível a outros utilizadores. Apesar de existirem protocolos e otimizações que possibilitam a utilização de gDNA degradado e de baixa quantidade, a avaliação continua determinante para a escolha acertada do kit/protocolo a seguir. De realçar que, embora a quantidade seja importante, a qualidade não deve ser esquecida. O DNA obtido deve estar livre de inibidores da reação de polimerização em cadeia (PCR), como concentrações altas de EDTA e sais, devendo ser utilizado o buffer de eluição mais apropriado para o protocolo seguinte.

A avaliação da qualidade compreende três etapas fundamentais: avaliação da integridade do DNA (tamanho), avaliação da pureza do DNA (contaminações) e avaliação da concentração do DNA (quantidade). Abaixo descrevem-se algumas metodologias que podem ser aplicadas para a análise destas três etapas.

A. Eletroforese em gel

Um dos métodos mais populares para a avaliação da integridade é a eletroforese horizontal em gel de Agarose a 1%. Um gDNA com boa qualidade apresenta um fragmento único, bem definido, com um tamanho >30 kb. Um DNA degradado apresenta vários fragmentos de menor tamanho, difundidos ao longo do gel.

A eletroforese em gel permite também avaliar a pureza no que respeita à presença de ácido ribonucleico (RNA), observada pela presença de uma banda difusa no fundo do gel. No caso de ser detetada a presença de RNA, deve ser realizado um tratamento com RNase antes das etapas seguintes.

Em alternativa, pode ser utilizada a eletroforese em gel de poliacrilamida (PAGE) ou equipamentos de eletroforese automática com kits apropriados para avaliação de gDNA. Alguns destes equipamentos calculam índices de qualidade, através de algoritmos, que permitem uma avaliação independente da interpretação do utilizador.

B. Espectrofotometria

A técnica de espectrofotometria permite determinar a pureza do gDNA a partir do cálculo de determinados comprimentos de onda (230, 260 e 280 nm). O rácio A260/A280 pode ser utilizado para indicar a pureza do DNA em relação a contaminantes como proteínas, que absorvem próximo dos 280 nm. O gDNA puro apresenta um rácio A260/A280 de 1.8–1.9, e o contaminado terá valores mais baixos. O rácio A260/A230 fornece uma estimativa da pureza do DNA em relação a contaminantes orgânicos como o Trizol, Fenol e Guanidina HCL, que absorvem próximo dos 230 nm. O gDNA puro apresenta um rácio A260/A230 de 2.0-2.2, enquanto o contaminado com estes produtos será mais baixo.

Este método permite também estimar a concentração de ácidos nucleicos presentes na amostra, a partir da leitura da absorbância a 260 nm, ainda que com menor sensibilidade que o método fluorométrico.

C. Fluorometria

O método fluorométrico é o mais utilizado para medir a concentração de gDNA de amostras a sequenciar por NGS. Este método, altamente seletivo, permite determinar a concentração de DNA de cadeia dupla, sem interferência do DNA de cadeia simples ou RNA. Mesmo na presença de contaminantes, é um método preciso e sensível, permitindo a quantificação de amostras mesmo com baixa concentração.

4.3. Armazenamento do gDNA

O gDNA extraído e purificado deve ser eluído num tampão adequado, como o tris-EDTA (TE), e armazenado em tubos de plástico hidrofóbico (*low binding*), como o polipropileno, testados para a ausência de inibidores de PCR, DNA, DNase e RNase, e de preferência, com fecho de rosca e vedação de borracha, de forma a evitar a evaporação. Embora o DNA se mantenha estável durante algumas semanas à temperatura ambiente, este deve ser conservado abaixo dos 0°C para minimizar a atividade das DNases. O congelamento a -20°C ou a -70/80°C são as opções mais indicadas para armazenamento a longo prazo.

Os ciclos de descongelamento-congelamento devem ser evitados. Caso seja necessário aceder com frequência ao DNA armazenado, é recomendada a divisão do DNA eluído em várias alíquotas, logo após a sua extração.

O correto armazenamento do DNA previne a degradação do mesmo. Contudo, após um longo período de armazenamento, é aconselhável a realização de um novo controlo de qualidade (QC) ao DNA.

5. Preparação de bibliotecas

De uma forma genérica, uma biblioteca de sequenciação é um conjunto de fragmentos da região a sequenciar, flanqueados por oligonucleotídeos conhecidos, os adaptadores. Efetivamente, o derradeiro objetivo de criar uma biblioteca é preparar o DNA extraído num formato que seja adequado à sequenciação por NGS.

Podem ser utilizadas várias opções existentes no mercado, contudo, a escolha do kit de sequenciação afeta a qualidade dos resultados obtidos, pelo que, como já foi dito anteriormente, a escolha deve ser adequada às amostras (e organismos) a sequenciar e ao instrumento de sequenciação que será utilizado (marca e modelo).

Durante a preparação das bibliotecas devem ser respeitadas as regras básicas de segurança e as boas práticas de manuseamento de amostras, reagentes e consumíveis, que acautelem a possibilidade de contaminação. É obrigatória a utilização de luvas e do EPI base. As luvas devem ser mudadas regularmente e sempre que houver suspeita de contacto com superfícies potencialmente contaminadas. Devem ser utilizadas pontas com filtro resistente a aerossóis e o seu comprimento deve ser tido em conta. As pontas curtas apenas devem ser utilizadas para pipetagens em placas e tubos baixos (por exemplo, tubos de 0.2 mL). Para as pipetagens em tubos e placas altas devem ser utilizadas pontas longas, de forma a ser evitada a contaminação cruzada provocada pelo toque accidental da haste da pipeta dentro dos recipientes. As áreas de trabalho e os instrumentos reutilizáveis devem ser descontaminadas diariamente,

de preferência, antes e após o procedimento. Todo o material utilizado deve estar estéril, livre de inibidores de PCR, DNA, DNase e Rnase.

No caso dos protocolos de construção de bibliotecas que usem amplificação, devem ser utilizadas duas áreas separadas de trabalho, uma de pré-PCR e outra de pós-PCR, unidirecionais, com equipamentos e instrumentos dedicados a cada área, como é o caso das micropipetas. O EPI também deve ser exclusivo de cada um dos espaços. Se possível, estes dois espaços devem estar em salas separadas, comunicando entre si por pequenas passagens para o material. Pelas mesmas razões, o armazenamento de DNA pré-amplificado e amplificado deve ser realizado em espaços distintos. Sempre que possível, devem também ser utilizados reagentes dedicados a cada espaço. Esta organização nem sempre é fácil quando são utilizados kits comerciais, onde apenas é disponibilizada uma alíquota de reagentes, mesmo que sejam comuns às duas áreas. Nesses casos, antes do início do procedimento, pode ser realizada a alíquotagem e separação do reagente em causa.

5.1. Metodologias

Na sequenciação de sequências curtas, destacam-se três tipos de protocolos: sequenciação completa do genoma (WGS), sequenciação completa do exoma (WES) e sequenciação dirigida (target), por exemplo, para painéis pré-definidos de genes de interesse. As sequenciações dirigidas podem ser realizadas recorrendo a dois métodos principais, sequenciação baseada em amplicons, onde os fragmentos são previamente selecionados e amplificados por PCR, ou sequenciação baseada em hibridização e captura, com sondas desenhadas para a região alvo. A escolha entre estes dois métodos deve ser determinada pela análise das vantagens e desvantagens de cada método.

O método baseado em amplicons tem como vantagem ser uma técnica mais simples e rápida, com um custo geralmente mais baixo por amostra e melhor taxa *on-target*, embora, possa comprometer a uniformidade.

O método de hibridização e captura possui uma maior sensibilidade, reprodutibilidade e possibilita a sequenciação simultânea de um maior número de regiões de interesse por painel (virtualmente ilimitado),

como é o caso da sequenciação do exoma completo ou do exoma clínico. Adicionalmente, esta opção resulta numa cobertura de sequenciação mais uniforme, maior complexidade e uma taxa mais baixa de falsos positivos.

Em geral, o método de hibridização e captura é o mais aconselhado para a sequenciação de grandes regiões de interesse e identificação de variantes raras. Em contraste, o método baseado em amplicons é recomendado para a sequenciação de alvos mais pequenos e para a identificação de variantes previamente conhecidas, como variantes de um único nucleótido (SNVs) e inserções/deleções (indels).

5.2. Etapas do fluxo de trabalho

A grande diversidade de kits existentes no mercado, leva a que, atualmente, existam protocolos de preparação de amostras cada vez mais diversificados. Contudo, existem três passos comuns à generalidade dos protocolos de sequenciação de fragmentos curtos de DNA: Fragmentação, Ligação de adaptadores e Purificação. No caso da sequenciação dirigida, dependendo do método escolhido, acresce a etapa de amplificação da região alvo por PCR (método baseado em amplicons) ou de enriquecimento por hibridização da região alvo com sondas e remoção dos fragmentos não hibridizados (método de hibridização e captura).

A. Fragmentação

A sequenciação requer a fragmentação do gDNA em fragmentos aleatórios com tamanhos semelhantes. Esta pode ser realizada por diferentes métodos, sendo os físicos e os enzimáticos os mais utilizados.

A fragmentação física é geralmente realizada por ultrassonificadores (fragmentação acústica), onde o tamanho dos fragmentos gerados é controlado pela intensidade e duração da sonicação. Esta metodologia produz uma distribuição de tamanhos mais homogénea entre diferentes amostras, uma vez que a fragmentação enzimática é mais sensível a variações na concentração de DNA e dos reagentes da reação, bem como a alterações na duração e temperatura da incubação. Contudo, a fragmentação enzimática torna-se uma alternativa atrativa para os laboratórios que não possuam, ou não pretendam adquirir,

um ultrassonicador, e para efeitos de simplificação do protocolo. Atualmente, as opções enzimáticas existentes no mercado foram também sujeitas a otimização e melhorias. Como forma de redução dos custos, e do tempo total do procedimento, alguns protocolos utilizam uma metodologia combinada de fragmentação e adição de uma porção de adaptadores (tagmentação), ou ainda a combinação destas duas etapas com a normalização da concentração de DNA na solução (transposomas ligados a beads).

B. Ligação de adaptadores

Os adaptadores são oligonucleotídeos sintéticos incorporados nas extremidades da sequência alvos e têm três funções principais na sequenciação: ligação à superfície de sequenciação (*flow cell*), ligação dos primers de sequenciação e multiplexação de amostras.

Na sequenciação, a biblioteca final é injetada nos canais da *flow cell*, que está revestida por oligonucleotídeos complementares aos adaptadores adicionados (adaptadores P5 e P7). Desta forma, os fragmentos unem-se à superfície onde ocorre a sequenciação. Contudo, a síntese da cadeia a sequenciar pelas polimerases só se inicia após a ligação do primer de sequenciação ao adaptador complementar introduzido. Qualquer falha durante a ligação destes adaptadores impossibilita a sequenciação.

Por último, podem ainda ser utilizados adaptadores para multiplexação de amostras. Esta estratégia permite a sequenciação simultânea de um grande número de bibliotecas numa mesma corrida de sequenciação. Estes adaptadores, conhecidos como indexes, são sequenciados juntamente com os fragmentos, possibilitando a identificação e agrupamento dos fragmentos provenientes de cada amostra, numa análise após a sequenciação. Porém, devem ser respeitadas as recomendações para a escolha dos indexes apropriados, e que vêm incluídas nos protocolos de preparação de bibliotecas, nomeadamente no que respeita à combinação ideal de indexes numa corrida. Estas recomendações são particularmente importantes quando são multiplexadas um pequeno número de bibliotecas, uma vez que a escolha dos índices não pode ser aleatória. Para mais informações

pode ser consultado, por exemplo, o guia para utilização de índices na Illumina (*Index Adapters Pooling Guide*, Illumina) [24]

C. Purificação

A purificação é a etapa mais comum nos protocolos de NGS e permite a remoção de material indesejado que pode interferir com as etapas a jusante, incluindo a própria sequenciação. Esta metodologia permite também a seleção de fragmentos por tamanho.

Existem dois métodos principais para a purificação e seleção de tamanhos: baseada em esferas magnéticas (*beads*) e eletroforese em gel. A utilização de *beads* é o método mais utilizado, porque é facilmente incorporado na rotina laboratorial, seja no processamento manual ou automatizado, e sendo frequentemente realizada em várias fases de um mesmo protocolo de preparação de bibliotecas. A variação da concentração de polietilenoglicol (PEG) e sais numa solução, provoca a ligação reversível das *beads* aos fragmentos de DNA. Com o recurso a um ímã, o complexo *beads*-DNA pode ser afastado da solução, para que seja possível a remoção do sobrenadante, lavagem do recetáculo e ressuspensão final do DNA.

A eletroforese em gel é uma opção mais exaustiva e demorada, contudo, é particularmente importante quando os fragmentos a isolar têm um tamanho próximo do dos fragmentos indesejados. Após a separação por tamanho a banda desejada é isolada e o DNA ressuspendido.

D. PCR

O PCR é uma etapa opcional, mas ainda realizada em grande parte dos protocolos de preparação de bibliotecas, principalmente quando é utilizada uma quantidade reduzida de gDNA, já que permite a sua amplificação.

5.3. Controlo de qualidade das bibliotecas

O CQ da biblioteca obtida é fundamental para a avaliação do sucesso do procedimento e antecipa o próprio sucesso da sequenciação. Os critérios de aceitação definidos pelo kit/protocolo utilizado devem ser

respeitados. Mais uma vez, se o laboratório, por experiência, alterar estes critérios, as alterações devem ficar registadas num PO acessível a outros utilizadores.

Tal como na avaliação das amostras de DNA, esta etapa compreende a avaliação da integridade da biblioteca (tamanho) e da sua concentração (quantidade). Contudo, é requerida uma avaliação ainda mais restrita. A avaliação do tamanho deve ser realizada por eletroforese automática e a quantificação pode ser realizada por fluorometria ou PCR quantitativo (qPCR). Estes métodos fornecem os dados mais precisos sobre a qualidade da biblioteca. Apenas devem ser sequenciadas as bibliotecas que passem nos critérios definidos, fornecidos pelo fabricante.

5.4. Desafios na preparação de bibliotecas

Apesar dos avanços na preparação das bibliotecas, o fluxo de trabalho continua, em muitos casos, trabalhoso, demorado e mesmo desafiante, especialmente quando se trabalha com amostras degradadas ou com baixas concentrações. A realização adequada do CQ ao gDNA e às bibliotecas torna possível controlar o fluxo de trabalho, contudo, este controlo não é total e a monitorização intermédia nem sempre é possível ou economicamente viável.

Os laboratórios devem estar preparados para detetar e solucionar os desafios que venham a surgir. A rede de distribuidores e empresas comerciais auxiliam os laboratórios nesta tarefa e, por vezes, os próprios protocolos incluem dicas rápidas para a resolução dos problemas mais comuns durante a preparação das bibliotecas. Contudo, o conhecimento prévio proporciona decisões e resoluções rápidas e eficientes, sem a necessidade e dependência de terceiros. Abaixo, descrevem-se alguns dos desafios mais frequentemente encontrados e respetivas formas de resolução.

A. Tamanho inadequado

O tamanho final dos fragmentos da biblioteca tem um grande impacto na eficiência e qualidade dos resultados de sequenciação, devendo ser respeitado com rigor o intervalo de fragmentos indicado

no protocolo. A obtenção de um tamanho de biblioteca inadequado geralmente ocorre por uma fragmentação ineficiente que leva à sobrefragmentação ou subfragmentação do DNA.

A sobrefragmentação pode levar à perda substancial do rendimento das bibliotecas durante a purificação, por apresentarem um tamanho mais baixo do que o esperado. Por vezes, é resultado da utilização de amostras degradadas, com tamanhos mais pequenos do que o expectável. Quando são utilizados amplicons, devem ser seguidas as recomendações do protocolo no que respeita ao tamanho destes fragmentos.

A subfragmentação provoca o aumento do tamanho final da biblioteca podendo originar uma sequenciação ineficaz. Os fragmentos de grande tamanho não se ligam tão eficazmente à *flow cell*. Existem duas causas comuns para a subfragmentação: a utilização de demasiado gDNA e a presença de inibidores enzimáticos durante a reação de fragmentação enzimática, como o EDTA ou sais. Este problema pode ser resolvido com a inclusão de uma etapa extra de purificação do gDNA após a sua extração.

As causas mais comuns da sobre- ou subfragmentação podem, desta forma, ser evitadas, com a realização de um CQ rigoroso do gDNA, realizado imediatamente antes da preparação das bibliotecas, de forma a ser descartada uma potencial degradação durante o armazenamento das amostras. Caso seja necessário, a etapa de fragmentação pode ser otimizada de acordo com a qualidade das amostras utilizadas. Por exemplo, na fragmentação enzimática, pode ser ajustada a duração e a temperatura da reação.

B. Contaminação

A presença de contaminantes na sequenciação pode levar à produção de ruído nas leituras obtidas e até a falsos achados. Como já foi mencionado anteriormente, devem ser cumpridas todas as boas práticas de manuseamento de amostras, reagentes e consumíveis que acautelam a possibilidade de contaminação.

Os fragmentos de DNA previamente amplificados noutros procedimentos/reações (amplicons) são a principal fonte de contaminação,

sendo especialmente prejudiciais nas fases de pré-amplificação. Alguns cuidados adicionais devem ser tomados durante a manipulação das amostras a montante e durante a preparação da reação de amplificação. Além de ser recomendada a utilização de uma área dedicada ao processamento pré-PCR, esta, se possível, deve ser dividida em duas subáreas, uma dedicada à preparação dos reagentes e outra para manipulação das amostras. Adicionalmente, a monitorização da contaminação, nos protocolos que incluam etapas de amplificação, deve ser realizada pela utilização de controlos negativos (NTC).

A contaminação cruzada, entre amostras, pode ser outra causa de contaminação. Devem ser utilizadas pontas com filtro e de tamanho adequado, para evitar que os líquidos ou aerossóis entrem em contacto com a haste da pipeta, e estas devem ser sempre trocadas entre cada pipetagem. Antes da abertura dos tubos, as amostras devem ser centrifugadas, assegurando que todo o volume se encontra depositado no fundo dos mesmos.

A contaminação pelos índices é uma forma adicional de contaminação, pelo que a manipulação dos índices deve ser realizada com especial atenção. Por exemplo, quando são utilizados tubos individuais destes adaptadores, deve ser aberto um tubo de cada vez, de forma a ser evitada a troca acidental das tampas.

C. Dímeros de adaptadores

Em amostras de baixa concentração ou qualidade, a etapa de ligação dos adaptadores deve ser realizada com precaução, nomeadamente no que respeita à concentração dos adaptadores na reação. Concentrações demasiado elevadas podem provocar a formação de dímeros. Ao contrário dos dímeros de primers, os dímeros de adaptadores ligam-se à *flow cell* e geram dados de sequenciação, levando à subtração da capacidade total da sequenciação, ou seja, a uma perda significativa de eficiência. Por se tratarem de fragmentos mais pequenos, podem-se ligar mais eficazmente do que as próprias bibliotecas pretendidas. A presença destes dímeros

compromete ainda a qualidade da sequenciação e, em casos extremos, pode levar à interrupção prematura da corrida, pela baixa diversidade apresentada (explicada nos tópicos seguintes). Apesar de ser um problema comum a todas as plataformas de sequenciação, as *flow cells* padronizadas, são ainda mais suscetíveis a este problema, resultando num maior impacto no desempenho da corrida.

Os dímeros de adaptadores são detetados durante o CQ final das bibliotecas (eletroforese automática), representados por um pequeno pico de fragmentos (geralmente entre 120-170 bp) inferior ao tamanho da biblioteca pretendida. Caso estejam presentes, deve ser realizada a otimização da etapa de ligação, por exemplo, com a modificação das concentrações utilizadas, aumentando a concentração inicial de DNA (preferencialmente) ou reduzindo a concentração de adaptadores na reação. A sua eliminação pode também ser alcançada com a realização de um passo de purificação adicional por *beads* ou por eletroforese em gel. Contudo, ambos resultam na redução do rendimento final, visível pela diminuição da concentração final das bibliotecas. Uma purificação inadequada também pode ser responsável pela presença de dímeros de adaptadores.

D. Duplicados

Devem ser seguidas com atenção as recomendações do protocolo para as etapas de amplificação, especificamente o número de ciclos recomendado para a quantidade de DNA utilizada. Contudo, poderá também ser necessária a otimização desta etapa. Este passo é suscetível à introdução de um viés na sequenciação, a criação de duplicados de PCR. Ou seja, a existência de várias cópias do mesmo fragmento, que podem levar a uma cobertura de sequenciação pouco uniforme.

A taxa de duplicados pode ser estimada durante a análise dos resultados da sequenciação. Quando a otimização da etapa não é suficiente podem ser utilizadas ferramentas bioinformáticas para a eliminação dos duplicados.

6. Sequenciação

A etapa de sequenciação ocorre, maioritariamente, de forma automatizada, na plataforma de sequenciação, sendo apenas necessária uma rápida preparação da biblioteca final, realizada imediatamente antes da biblioteca ser carregada no equipamento. A preparação é realizada segundo protocolo disponibilizado pelo fornecedor, específico para cada plataforma. Este processo envolve a junção das bibliotecas num único tubo (caso seja utilizada uma reação em multiplex e se este passo não foi realizado durante a preparação das bibliotecas), desnaturação com hidróxido de sódio (NaOH) e diluição. A concentração final utilizada é um fator crítico para o desempenho da sequenciação. Devem ser respeitadas as recomendações do fabricante, contudo, a otimização é geralmente indispensável.

Nesta fase, pode ser adicionado à biblioteca final um controlo de sequenciação, eg. Phix, preparado segundo o mesmo protocolo das bibliotecas. Este controlo consiste numa biblioteca comercial, não indexada, pronta a utilizar, de um genoma bacteriófago bem caracterizado. A adição do controlo de corrida, tipicamente utilizado a $\geq 1\%$, é recomendada para a monitorização da sequenciação de uma corrida Illumina, incluindo a geração de clusters e a sequenciação. Este controlo também é utilizado na validação dos equipamentos Illumina, correndo sozinho numa sequenciação de teste.

6.1. Etapas da sequenciação

Existem dois passos principais na química de sequenciação por síntese (SBS): Geração de clusters e Sequenciação. Após a ligação dos fragmentos à *flow cell*, cada fragmento é amplificado em clusters distintos (geração de clusters), possibilitando a deteção do sinal durante a sequenciação propriamente dita. Durante a sequenciação, são incorporados trifosfatos de desoxirribonucleótidos marcados com fluorescência, por uma polimerase, durante ciclos sequenciais de síntese das cadeias complementares de cada fragmento. Em cada adição/ciclo, é registada a fluorescência emitida (por análise de imagem). O processo ocorre em

paralelo em todos os clusters gerados. O número final de leituras produzidas e o seu tamanho varia consoante a plataforma de sequenciação e o kit utilizado.

6.2. Monitorização da corrida

A monitorização da corrida ocorre a partir da interpretação de parâmetros gerados ao longo da sequenciação. Algumas métricas de qualidade e eficiência da corrida, podem ser acompanhadas mesmo antes da sua conclusão. Determinados parâmetros podem ser visualizados no software do próprio equipamento, outros estão acessíveis, por exemplo, através de outros softwares/servidores como o Illumina BaseSpace. De entre as métricas geradas, as cinco mais frequentemente analisadas são: Rendimento (Gb), Qualidade Q30 (%), Taxa de Erro (%), Densidade dos clusters (K/mm²) e Percentagem de clusters que passam os filtros de qualidade-PF (%). Apesar de representadas de forma independente, estas métricas, correlacionam-se, devendo ser interpretadas em conjunto.

O rendimento representa a quantidade de dados gerados em Gigabases e a %Q30 a qualidade desses dados. A classificação Q (*Q-score*) é baseada na probabilidade de uma identificação incorreta das bases, sendo que o Q30 representa a classificação referência da qualidade, em que a probabilidade de identificação incorreta é de 1 em 1000 (precisão 99.9%). A métrica é apresentada como a percentagem de bases com uma classificação maior ou igual a Q30. Por sua vez, a taxa de erro reflete a percentagem de bases identificadas incorretamente, mas é calculado a partir do controlo Phix, pelo que só deve ser monitorizado quando este controlo é utilizado.

Entre todas as métricas, a densidade dos clusters (K/mm²) é a que tem um dos maiores impactos no desempenho da sequenciação. Quando é alcançada a densidade ideal é maximizada a qualidade e a capacidade de sequenciação, com a produção total de dados prevista. Desta forma, a densidade é uma medida proporcional à qualidade dos dados e deve ser avaliada em conjunto com a percentagem de clusters que passam os filtros de qualidade-PF (%). Os clusters que não passam a filtragem são removidos da análise. Densidades demasiado elevadas resultam na diminuição da %PF e %Q30 e, por sua vez, do rendimento, mas densidades

demasiado baixas também resultam na diminuição do rendimento, como será demonstrado no tópico seguinte.

6.3. Desafios na sequenciação

Tal como na preparação de bibliotecas, a sequenciação acarreta alguns desafios que requerem reflexão. Seguidamente, são descritos dois dos desafios mais frequentes e respetivas formas de resolução.

A. Baixa diversidade

A diversidade dos nucleótidos presentes é fundamental para o sucesso da sequenciação. Uma biblioteca diversa possui proporções semelhantes dos quatro nucleótidos em cada ciclo da sequenciação, caso a percentagem de GC ronde os 50%. Esta deve se mantida tanto na sequenciação do fragmento alvo como na dos indexes, que, dada a sua natureza, já são caracterizados por uma menor diversidade. Por este motivo, como foi referido anteriormente, a escolha dos indexes não pode ser aleatória, principalmente quando são multiplexadas um baixo número de bibliotecas. Sempre que vários clusters emitem o mesmo sinal em cada ciclo, o detetor pode ter dificuldade em identificar os clusters emissores.

A solução mais simples para se sequenciar uma biblioteca pouco diversa é a incorporação de diversidade através da utilização de Phix. O Phix possui uma composição equilibrada de bases (aproximadamente 45% GC e 55% AT) e, quando sequenciado, produz a emissão balanceada de sinais em cada ciclo. Neste contexto, deve ser aumentada a proporção de Phix na sequenciação, em bibliotecas previsivelmente pouco diversas. No mínimo, é recomendado pela Illumina a utilização de 5-10% de Phix (embora varie consoante a plataforma), mas encontrar a proporção ideal, para cada caso, requer otimização.

B. Geração incorrecta de clusters

Na prática, a sequenciação deve ser realizada com uma densidade de clusters suficientemente elevada para maximizar a produção total de dados, mas razoavelmente baixa para preservar a qualidade pretendida.

Quando há um excesso da concentração, a densidade de *clusters* torna-se demasiado elevada, originando problemas na análise das imagens adquiridas, incluindo problemas no foco e na identificação dos clusters. Isto, em casos extremos, pode mesmo levar à interrupção prematura da corrida. As *flow cells* padronizadas não são tão suscetíveis a este problema, nem ao cenário anteriormente apresentado, i.e. a presença de bibliotecas com baixa diversidade. Por outro lado, densidades demasiado baixas não levam a problemas de reconhecimento, mas sim a uma diminuição do rendimento, com menos dados produzidos.

Estas situações são rapidamente identificadas, após os primeiros ciclos de sequenciação, pela métrica de percentagem de clusters gerados. Quando há excesso de densidade há também um conseqüente decréscimo do Q30 e %PF, e o rendimento da corrida também poderá ser afetado, com menos dados gerados. Adicionalmente, esta situação pode provocar a identificação imprecisa dos indexes, exacerbando a sua baixa diversidade.

Para a obtenção da densidade ideal é fundamental a realização de um controlo de qualidade e uma preparação das bibliotecas rigorosa antes da corrida. A desnaturação da biblioteca deve ser realizada com uma diluição fresca de NaOH de forma ser mantido o pH básico (pH > 12.5). Ao longo do tempo, o pH da solução de NaOH diluído diminui. Após a desnaturação deve ocorrer a neutralização do NaOH com Tris-HCl, de forma a que a presença do NaOH na solução final seja inferior a 1 mM. Quando algumas destas duas situações não é respeitada, ocorre a inibição da formação dos clusters.

7. Automatização de processos

Atualmente, estão ao dispor dos laboratórios uma elevada diversidade de equipamentos para procedimentos automatizados. Cabe ao laboratório definir as opções mais ajustadas à sua realidade, que se enquadram no modelo operacional praticado, não sendo objetivo deste manual sugerir ou aconselhar qualquer marca ou procedimento em específico.

Os laboratórios devem acompanhar os progressos e inovações no campo, e estar preparados para mudanças estratégicas para fazer face a novas realidades e necessidades.

A automatização tem vindo, erradamente, a ser vista como uma solução apenas para a sequenciação em larga escala. Contudo, a padronização de processos, fundamental na automatização, é uma estratégia importante em todos os níveis e setores, incluindo nos laboratórios mais pequenos, clínicos ou de investigação. A automatização, além de ser essencial para o aumento da produtividade dos laboratórios e diminuição do tempo de resposta e dos custos finais, também se traduz na maior uniformização e reprodutibilidade dos resultados alcançados, e mesmo na redução da taxa de erro, fatores importantes em qualquer escala laboratorial.

Adicionalmente, a automatização pode ser introduzida em todas as fases (pré-analítica, analítica e pós-analítica), não acomodando apenas o recurso de robots e pipetadores automáticos. O seu grande objetivo é a utilização de técnicas que tornem um procedimento mais eficiente, seja pela automatização de etapas ou pela automação integral de um laboratório. Automatizar etapas pode ser alcançado, por exemplo, com a simples utilização de códigos de barras, ou simplesmente com a eliminação ou simplificação de atividades ou documentação desnecessária. Cabe aos laboratórios definirem as estratégias que melhor se enquadram com o seu fluxo operacional e os seus objetivos.

ANÁLISE E GESTÃO DE DADOS

Apesar do NGS já ser uma tecnologia bem estabelecida em muitos laboratórios clínicos e de investigação, continua a exigir conhecimentos muito especializados relativamente à prática laboratorial, mas também à bioinformática.

Após a sequenciação, os dados são convertidos e disponibilizados em ficheiros FASTQ que, entre outros dados, incluem as leituras obtidas. Para a análise dos dados podem ser utilizadas, de forma sequencial, ferramentas comerciais ou desenvolvidas internamente (*pipelines*).

Os tópicos seguintes incluem algumas etapas do processamento bioinformático dos dados, recomendações para o desenvolvimento de *pipelines* bioinformáticas de qualidade e recomendações para a análise dos dados. Inclui ainda os principais desafios, incluindo no que respeita aos requisitos técnicos e ao armazenamento de toda a informação gerada. Como referido anteriormente, a tecnologia de sequenciação da Illumina é a mais utilizada na Região Centro, pelo que os tópicos seguintes serão desenvolvidos no contexto desta tecnologia.

1. Requisitos técnicos: software & hardware

A escolha dos softwares a utilizar é um dos primeiros desafios, já que estes devem ser adequados à análise pretendida, permitindo explorar

integralmente os dados gerados. Adicionalmente, são necessários recursos de hardware adequados para processar os dados gerados pelo sequenciador de forma eficiente. São processos computacionalmente intensivos sendo recomendado um processador (CPU) com vários núcleos e uma frequência de *clock* (GHz) alta para lidar com o processamento paralelo e acelerar o tempo de execução. A quantidade de memória (RAM) disponível é importante para acomodar os dados de sequenciação e executar todos os algoritmos do *pipeline* bioinformático. Recomenda-se uma quantidade suficiente para armazenar os arquivos de sequenciação e executar os programas de análise sem sobrecarregar o sistema. Geralmente, recomenda-se uma configuração mínima de 128 GB, apesar de quanto mais RAM, melhor. Embora não seja estritamente necessário, uma placa gráfica (GPU) pode acelerar certos algoritmos usados em algumas etapas do *pipeline* bioinformático, nomeadamente no alinhamento e identificação das variantes. Estas ferramentas podem ser otimizadas para aproveitar o poder de processamento paralelo das GPUs, o que resulta num processamento mais rápido dos dados. No entanto, nem todos os softwares de análise de NGS suportam o uso de GPUs. Muitas ferramentas e *pipelines* de análise de NGS são desenvolvidas e otimizadas para funcionar em sistemas Unix/Linux, pela sua ampla compatibilidade com ferramentas de bioinformática populares, capacidade de automação de tarefas utilizando *scripts* e estabilidade no processamento de dados em larga escala. De notar, que as necessidades de hardware podem variar dependendo da escala do projeto e da análise realizada. É sempre recomendável consultar atempadamente as especificações e requisitos de hardware, softwares e algoritmos a utilizar, para garantir uma configuração adequada do sistema.

2. Processamento de dados

O processamento bioinformático dos dados para a deteção de alterações genéticas é geralmente específico da metodologia utilizada e customizado de acordo com as necessidades de cada laboratório. Apesar de existirem

no mercado várias opções para a análise e deteção de variantes, muitos laboratórios optam pela construção das suas próprias *pipelines* bioinformáticas. No entanto, existem passos de análise que são transversais às duas abordagens. O pipeline bioinformático (Figura 2) incorpora uma série de processos computacionais de análise de dados, que culmina na obtenção de uma lista de variantes genómicas prontas a serem interpretadas/classificadas. Podem ainda ser adicionados à *pipeline* filtros customizados para a filtragem das variantes mais relevantes ou eliminação das menos relevantes (por exemplo, filtros por fenótipo, frequência alélica ou qualidade).

2.1. Visão geral do pipeline

O *pipeline* bioinformático típico é composto por várias etapas interconectadas, incluindo pré-processamento de dados, alinhamento, identificação de variantes, anotação de variantes e análise posterior. Cada etapa desempenha um papel crucial no fluxo geral de análise. As secções seguintes detalham os procedimentos recomendados para cada etapa.

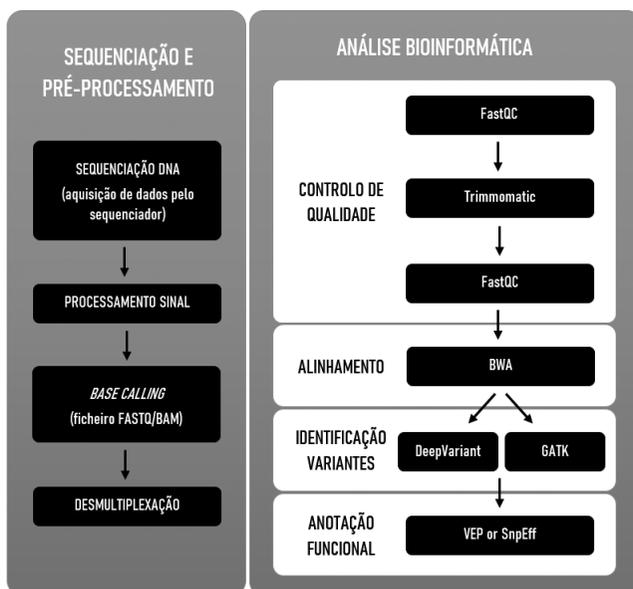


Figura 2 – Etapas do *pipeline* de análise bioinformática para NGS.

2.2. Pré-processamento

O primeiro passo tem como objetivo a avaliação de qualidade e pré-processamento dos dados sequenciados, visando garantir a confiabilidade e precisão das análises bioinformáticas subsequentes. Nesta etapa, os ficheiros com dados não tratados, gerados pelo sequenciador (*binary base cal*, BCL), são convertidos em formato FASTQ usando softwares específicos, como o *bcl2fastq* (Illumina). Estes ficheiros compilam informações referentes à amostra, incluindo a identificação da amostra, sequências dos adaptadores para desmultiplexação e uma medida de qualidade (*Q-score*) para cada base.

A sequenciação multiplexada tornou-se amplamente adotada, permitindo a sequenciação simultânea de várias amostras numa única corrida. Como foi referido anteriormente, envolve a marcação ou indexação de cada amostra, maximizando a eficiência e o rendimento da sequenciação, ao permitir a análise simultânea de um grande número de amostras. A desmultiplexação separa as leituras de sequenciação agrupadas nas suas respetivas amostras com base nos identificadores únicos adicionados durante a multiplexação. Após a desmultiplexação, os ficheiros FASTQ estão prontos para o CQ.

O CQ envolve várias etapas, incluindo a avaliação de métricas de qualidade dos dados, a deteção de contaminação por adaptadores e a remoção de leituras de baixa qualidade. Existem várias ferramentas disponíveis para avaliar as diferentes métricas de qualidade dos dados de NGS, como por exemplo o FASTQC (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). O FASTQC recebe dados brutos de sequenciação no formato FASTQ e gera um relatório detalhado de qualidade geral dos dados, com estatísticas resumidas para cada métrica analisada, como a profundidade de sequenciação, qualidade das bases sequenciadas, conteúdo de GC e contaminação por adaptadores. Isso permite uma avaliação rápida da qualidade geral dos dados e a identificação de potenciais problemas que possam afetar a precisão e confiabilidade da análise subsequente.

Ferramentas como o Trimmomatic [25] são posteriormente usadas para remover leituras de baixa qualidade, adaptadores e artefactos.

São aplicados limites de qualidade e filtros de comprimento para garantir a integridade dos dados. Por exemplo, uma classificação Q20 é normalmente usada como limite e indica uma probabilidade de 1% de uma identificação incorreta da base (precisão 99%). Medidas mais altas correspondem a probabilidades de erro mais baixas e maior precisão na identificação de base, como foi referido no Capítulo 1. O Trimmomatic oferece vários módulos e funcionalidades que podem ser personalizados para atender a protocolos específicos de sequenciação e características dos dados, desempenhando um papel vital no pré-processamento e controle de qualidade de dados de NGS. Algumas características e funções-chave do Trimmomatic incluem:

A. Remoção de adaptadores

Identifica e remove sequências de adaptadores que são frequentemente introduzidas durante a preparação da biblioteca. Estas sequências podem interferir na análise subsequente e afetar a precisão dos resultados.

B. Limites de qualidade

Os limites de qualidade I, permitem remover no início e no final da sequência bases que não tenham uma qualidade (classificação Q) superior ao limite definido pelo utilizador. Os limites de qualidade II, permitem remover as leituras de baixa qualidade, caracterizadas por Q baixos, e que podem afetar adversamente a precisão das análises subsequentes. O Trimmomatic implementa métodos de remoção de qualidade, como remoção por janela deslizante e remoção de base inicial/final, para remover bases de baixa qualidade ou leituras com qualidade geral baixa.

C. Remoção por comprimento

Remoção de leituras de sequenciação que ficam abaixo de um determinado limite de comprimento e que, por isso, podem não possuir informação suficiente para uma análise confiável. O Trimmomatic permite que os usuários especifiquem um comprimento mínimo

de leitura, e qualquer leitura abaixo desse limite é descartada do conjunto de dados.

D. ANÁLISE de leituras emparelhadas (*paired-end*)

O Trimmomatic é capaz de lidar com dados de sequenciação de leituras emparelhadas, em que as leituras são geradas a partir de ambas as extremidades de fragmentos de DNA. Esta ferramenta garante que as leituras emparelhadas sejam cortadas de forma consistente para manter o alinhamento e a integridade dos dados.

Em geral, o Trimmomatic ajuda a melhorar a confiabilidade e a precisão das análises subsequentes, e.g. alinhamento do genoma, identificação de variantes e análise de expressão genética, oferecendo ainda flexibilidade na escolha dos formatos de saída para as leituras processadas, como o formato FASTQ ou formatos compactados como GZIP, que ajuda a reduzir os requisitos de armazenamento.

2.3. Alinhamento

As sequências são alinhadas contra um genoma humano de referência, sendo atualmente os mais utilizados o *Genome Reference Consortium Human Build 37* (GRCh37, hg19), lançado em 2009, e o *Genome Reference Consortium Human Build 38* (GRCh38, hg38), lançado em 2013. Os algoritmos de alinhamento, como o Bowtie2 ou o BWA,[26] são utilizados para alinhar as sequências de todas as leituras ao genoma de referência. Nesta etapa são criados ficheiros de Mapeamento de Sequência (SAM), Binário-SAM (BAM) [27] ou comprimido em relação à referência CRAM [28], que representam os resultados do alinhamento. Diversas métricas de qualidade do alinhamento, incluindo profundidade e amplitude de cobertura, qualidade média de mapeamento e taxa de mapeamento, são calculadas para avaliar a qualidade dos alinhamentos e a confiabilidade dos dados. Estas métricas de qualidade são tipicamente calculadas e analisadas usando ferramentas bioinformáticas e softwares especificamente desenhados para a avaliação, como o Samtools [27] e o Picard (<http://broadinstitute.github.io/picard>).

2.4. Identificação de variantes de nucleotídeo único, inserção e deleção

Para identificar variantes de nucleotídeo único (SNVs) ou inserções/deleções curtas (indels), são aplicados algoritmos de identificação de variantes, como o GATK (<https://gatk.broadinstitute.org/hc/en-us>) ou o DeepVariant [29]. Estes algoritmos determinam a probabilidade de um determinado genótipo ocorrer, com base em índices de qualidade, cobertura de sequenciação e localização de bases. Recomenda-se correr ambos os algoritmos, GATK e DeepVariant, e apresentar os dois resultados ao operador. Devido à complexidade do procedimento, os algoritmos poderão originar resultados distintos, em zonas de maior complexidade. As alterações identificadas são reportadas em ficheiros com o formato VCF (<https://samtools.github.io/hts-specs/VCFv4.2.pdf>) contendo informações sobre as variantes, incluindo locais cromossómicos, bases de referência e alternativas, métricas estatísticas e genótipos para cada amostra. As ferramentas de filtragem, como a vcftools [30] e a SnpSift, podem ser posteriormente usadas para filtrar variantes com base em critérios específicos.

2.5. Identificação de variações no número de cópias

Nesta etapa, algoritmos de deteção de variações no número de cópias (CNV), como o VarScan2 (<https://varscan.sourceforge.net/>), o CNVKit (<https://cnvkit.readthedocs.io/en/stable/>), e o OncoCNV (<http://boevalab.inf.ethz.ch/ONCOCNV/>), são utilizados para identificar ganhos ou perdas de cópias. A cobertura é normalizada em relação às amostras controlo sem variantes conhecidas. São avaliadas as diferenças na cobertura de leituras em comparação com regiões circundantes e frequências alélicas de alelos alternativos. O resultado é a identificação e anotação de CNVs para análises posteriores. Todavia, a identificação de CNVs apenas é exequível quando se consegue sequenciar um número razoável de controlos.

2.6. Anotação e filtração de variantes

As variantes identificadas nas etapas anteriores são anotadas com informações sobre o gene, efeito da variante (por exemplo, *missense*, *nonsense*, *splice-site*), frequência e grau de associação a doenças. Bancos

de dados e ferramentas públicas podem ser usados para obter anotações funcionais e informação complementar. Adicionalmente, podem ser aplicados filtros para priorizar variantes com base em critérios predefinidos, como frequência alélica, patogenicidade e impacto funcional.

Esta tarefa pode ser realizada com o recurso a dois softwares, o SnpEff ou o Variant Effect Predictor (VEP). Bases de dados adicionais terão de ser alocadas aos softwares, caso o operador queira enriquecer as variantes com informação extra.

Os ficheiros de entrada e saída têm o formato VCF, especificado anteriormente. No entanto, recomenda-se a compressão dos ficheiros, estando os softwares de bioinformática preparados para consumir ficheiros comprimidos.

No caso específico do ficheiro VCF, sendo um formato tabular, recomenda-se o algoritmo de compressão block GZIP (bGZIP), pois efetua compressão por blocos de dados. Tem a vantagem de possibilitar a criação de um índice (tabix -p vcf <nome do ficheiro>) facilitando a procura de variantes através da coordenada posicional e da identificação do cromossoma. Se determinado ficheiro VCF for comprimido unicamente com o algoritmo GZIP, no caso de procura de uma variante específica, o algoritmo percorrerá todo o ficheiro até à sua correspondência.

2.7. Interpretação e classificação de variantes

Nesta etapa, é realizada a interpretação das variantes candidatas com base no seu impacto funcional, associações conhecidas com doenças e relevância para o fenótipo do paciente, culminando na sua classificação. As variantes são priorizadas com base na sua significância clínica e relevância para a condição do paciente.

A classificação de variantes é uma parte essencial do trabalho em Genética Médica e desempenha um papel crucial na interpretação dos resultados dos testes genéticos. A classificação da significância clínica das variantes genéticas deve ser realizada segundo as diretrizes internacionais descritas pelo American College of Medical Genetics (ACMG) e Association for Molecular Pathology (AMP) [31], e pelo Clinical Genome Resource (ClinGen). A classificação das variantes é feita em cinco

categorias, seguindo a terminologia *benigna, provavelmente benigna, de significado clínico desconhecido, provavelmente patogénica e patogénica*, e tem como base 28 critérios suportados por evidências, onde são atribuídos diferentes pesos, em função da informação disponível sobre uma determinada variante. As evidências podem incluir, por exemplo, dados populacionais, computacionais, funcionais ou de segregação.

Este processo auxilia os profissionais de saúde na tomada de decisões clínicas e no aconselhamento genético. Além disso, a transparência e a partilha de informações, tal como o recurso a bancos de dados, incluindo o ClinVar (<https://www.ncbi.nlm.nih.gov/clinvar/>), promovem a colaboração e o avanço do conhecimento em Genética. É importante ressaltar que as normas de classificação de variantes estão em constante evolução, à medida que novas evidências e descobertas científicas surgem. Portanto, é essencial que os profissionais da área se mantenham atualizados sobre as diretrizes mais recentes e participem de discussões e colaborações para aprimorar cada vez mais a classificação de variantes genéticas.

3. Armazenamento de dados

A sequenciação NGS gera uma enorme quantidade de dados que necessitam de ser guardados. Para além disso, tendencialmente, o desenvolvimento de novas tecnologias tem sido acompanhado por um aumento substancial da quantidade de dados gerados, o que tem aumentado a tensão dos laboratórios no que diz respeito aos custos associados com o seu armazenamento.

Sendo considerada uma limitação constante dos laboratórios, cada instituição deve definir a sua política de armazenamento. Na generalidade, aconselha-se o armazenamento durante, pelo menos, um ano. Contudo, num contexto de diagnóstico, e dado o custo de cada sequenciação, este prazo deve, sempre que possível, ser estendido. Esta política tem sido implementada em vários laboratórios. Por exemplo, alguns laboratórios, oferecem a possibilidade de reanálise dos dados gerados. Esta opção reduz a necessidade de repetição de todo o processamento

laboratorial da amostra e diminuí assim os custos inerentes ao estudo de cada indivíduo. Como exemplo, dá-se a realização de uma sequenciação do Exoma, uma só vez por indivíduo, com armazenamento dos ficheiros primários, o que permite a filtragem de genes em diferentes painéis de interesse, conforme solicitado pelo clínico. Esta opção só será disponibilizada caso o laboratório tenha capacidade de armazenamento de dados por um longo período. Caso contrário, haverá a necessidade de se repetir a sequenciação a cada pedido.

Num contexto de investigação, considera-se necessário o armazenamento dos dados até, pelo menos, à revisão por pares e publicação do artigo científico onde estes dados serão incluídos. De notar ainda, que os mesmos dados, caso haja consentimento para tal, podem ser utilizados em estudos de investigação adicionais, pelo que caso haja essa possibilidade, recomenda-se a conservação dos dados por um período mais longo de tempo.

De forma a prevenir a perda acidental dos dados, os laboratórios devem ainda possuir uma cópia de todos os dados gerados. Apesar desta política não estar acessível a todas as instituições, sempre que possível, os laboratórios devem praticá-la. Abaixo descrevem-se algumas possibilidades de armazenamento de dados que as instituições podem adotar.

A. armazenamento local

Armazenamento utilizando servidores locais ou discos de armazenamento externo. Os discos externos devem ser geridos com precaução, não devem ser transportados para fora da instituição, nem utilizados em computadores/equipamento facilmente expostos a vírus. Adicionalmente, a fragilidade destes sistemas pode levar ao comprometimento dos dados, pelo que não são a solução mais recomendada [35].

B. armazenamento em nuvem

Armazenamento utilizando os servidores em nuvem, disponibilizados por algumas instituições. Um destes exemplos é a Illumina, que fornece, juntamente com os seus equipamentos, acesso à ferramenta

BaseSpace. Esta ferramenta permite o armazenamento de 1 Tb de dados gratuitamente, podendo esta capacidade ser aumentada a qualquer momento através da compra de espaço adicional. Esta ferramenta também permite a utilização de aplicações desenvolvidas pela Illumina para a análise dos dados gerados, mediante o pagamento de uma taxa, ou a partilha facilitada dos dados entre os utilizadores desta ferramenta (gratuito).

(Página deixada propositadamente em branco)

QUALIDADE

A qualidade pode ser enquadrada no produto, serviço ou no próprio processo, com o objetivo final de satisfazer as necessidades e requisitos do público alvo, o cliente. No setor laboratorial, em que se enquadra o NGS, a qualidade centra-se nas diferentes fases que constituem o processo laboratorial (fase pré-analítica, analítica e pós-analítica) com o objetivo de garantir a conformidade dos processos e a melhoria contínua.

A qualidade deve ser preocupação constante dos laboratórios e dos seus colaboradores. A má prática, acarreta consequências económicas, sociais e práticas na vida dos pacientes, como a perda completa dos resultados, ou, mais grave, a troca de dados/amostras entre pacientes e o consequente erro de diagnóstico. Num processo de numerosos passos, como é o caso do NGS, são múltiplas as possibilidades de ocorrerem erros e nem todos são facilmente identificáveis. A falta de qualidade leva ainda à perda de confiança nas organizações e laboratórios envolvidos e, por fim, a perdas económicas. Apesar da implementação de um sistema de qualidade acarretar custos significativos, estes podem ser mais baixos que os custos inerentes a erros por más práticas laboratoriais. Adicionalmente, a otimização dos processos leva, a longo prazo, à redução dos custos.

1. Sistemas de gestão da qualidade

Gerir a qualidade num laboratório clínico ou de investigação é um processo contínuo e dinâmico, que requer a participação ativa de todos

os profissionais envolvidos no processo, e não deve basear-se na implementação de regras estanques e perpétuas. Os laboratórios devem nomear um responsável interno para garantia da qualidade, com formação adequada para esta tarefa.

A norma ISO 9001 (revista para 9001:15) é um bom ponto de partida para as instituições e laboratórios planearem, implementarem e avaliarem os seus procedimentos e sistema de qualidade. Esta norma, de carácter abrangente, segue uma abordagem *Plan-Do-Check-Act* (PDCA), também conhecida como ciclo de *Deming*, que se inicia com a organização e o planeamento dos processos e objetivos do sistema e culmina com a criação de ações de melhoria contínua. Os laboratórios, caso desejem, podem ser certificados para a norma ISO 9001, demonstrando, através de auditorias externas independentes, realizadas por organizações devidamente autorizadas, que o seu sistema de gestão está de acordo com esta norma. Contudo, esta norma é comum a diferentes áreas e concentra-se no controlo do processo.

Os laboratórios podem também implementar a acreditação ISO 15189 (revista para 15189:14), adaptada para os laboratórios clínicos, que reconhece a competência dos laboratórios em realizar determinadas tarefas [36]. Globalmente, está direcionada para a avaliação e manutenção de um programa de melhoria contínua do desempenho. Esta norma está dividida em requisitos Técnicos e de Gestão. De uma forma genérica, os requisitos técnicos incluem os recursos humanos, as instalações e os recursos tecnológicos. Os requisitos de gestão englobam o sistema de gestão da qualidade adaptado à norma 15189. De realçar a necessidade do controlo adequado dos documentos e registos do laboratório, controlo das não conformidades, ações corretivas e preventivas, e ainda a participação em auditorias (internas ou externas) ao laboratório.

A certificação é um processo obrigatório para um laboratório clínico. Já a acreditação é um processo voluntário em Portugal. Todavia, ambas devem ser vistas como modelos a alcançar na garantia da qualidade laboratorial, quer para os laboratórios clínicos, como para os não clínicos, como é o caso dos laboratórios de investigação na área da Genética e da Genómica. Os laboratórios que não prossigam com a certificação e/

ou acreditação podem, e devem, sempre que possível, trabalhar nestes moldes, garantindo a qualidade. Por outro lado, o cumprimento prévio da legislação facilitará o processo de integração aquando de uma posterior implementação.

Por outro lado, as unidades privadas que desenvolvam atividades laboratoriais no âmbito da Genética Médica, devem, obrigatoriamente, estar licenciadas para o efeito. A Portaria n.º 167/2014 de 21 de agosto [37], estabelece, para as unidades privadas, os requisitos mínimos relativos ao licenciamento, instalação, organização e funcionamento, recursos humanos e instalações técnicas, englobando os respetivos postos de colheitas. Contudo, mais recentemente, pelo Despacho n.º 14603/2022, de 21 de dezembro [38], foi criado um grupo de trabalho com a missão de estudar e propor os requisitos técnicos comuns a todos os estabelecimentos prestadores de cuidados de saúde com vista ao licenciamento, onde se inclui a tipologia dos laboratórios de genética médica e respetivos postos de colheita, dando continuidade ao trabalho desenvolvido pelo grupo de trabalho criado por Despacho n.º 14174/2016, de 25 de novembro [39], que apenas levou à publicação da Portaria n.º 392/2019, de 5 de novembro [40], referente aos laboratórios de patologia clínica e respetivos postos de colheita.

2. Critérios de aceitação e rejeição

Como foi sendo referido ao longo deste documento, os laboratórios devem definir os seus critérios para aceitar e rejeitar amostras, de acordo com cada procedimento. A definição destas regras otimiza o desempenho dos laboratórios, garante a qualidade do serviço e minimiza a possibilidade de erros associados à inadequação das amostras.

Os critérios de aceitação podem incluir a apresentação das amostras, por exemplo, definir os dados de identificação que devem acompanhar cada amostra ou o tubo que deve ser utilizado para a colheita, mas também podem referir-se a critérios quantificáveis, como volumes mínimos necessários ou mesmo concentrações.

Tanto os critérios de aceitação como os de rejeição devem ser mencionados e detalhados nos POs de cada aplicação.

3. Programas de controlo de qualidade

A implementação/participação em programas de controlo de qualidade internos e externos, é a forma mais simples de se avaliar e assegurar a qualidade laboratorial, bem como identificar as possíveis falhas e oportunidades de melhoria.

3.1. Controlo de qualidade interno

O controlo de qualidade interno (IQA), permite avaliar a qualidade regular da execução laboratorial. Este pode ser realizado, por exemplo, pela monitorização dos resultados, através da utilização de amostras/controles internos ou comerciais com valores/resultados previamente conhecidos. A utilização destas amostras controlo permite avaliar a precisão das sequenciações realizadas, a reprodutibilidade e, ao mesmo tempo, detetar e corrigir rapidamente possíveis erros. Permitem ainda a validação de novas técnicas, reagentes e equipamentos. Para a utilização desta metodologia os laboratórios devem previamente especificar os limites de aceitação dos resultados obtidos pelos controlos.

A realização de auditorias internas auxilia no cumprimento dos sistemas de gestão da qualidade do laboratório, sendo uma boa prática voluntária dos laboratórios. Porém, as auditorias internas têm carácter obrigatório e anual para os laboratórios acreditados pela ISO 15189.

O laboratório deve ainda manter uma rotina de formação contínua do pessoal técnico envolvido, com ações corretivas e documentação atualizada sobre as metodologias aplicadas, incluindo a frequência com que são realizadas e os resultados alcançados.

3.2. Controlo de qualidade externo

Os ensaios externos de qualidade (EQA) são sistemas de avaliação do desempenho de um laboratório através de uma agência externa. Apesar de

existirem vários métodos, os ensaios de proficiência são os mais utilizados. Nestes ensaios, são enviadas para os laboratórios amostras cujos resultados são por eles desconhecidos. Estas devem ser incorporadas na rotina laboratorial, processadas e tratadas sem distinção, utilizando os mesmos métodos, equipamentos e meios técnicos envolvidos no processamento habitual do laboratório. O objetivo é mimetizar o processamento de rotina. A partilha de resultados entre laboratórios participantes é proibida [41]. Após a análise, os resultados obtidos devem ser enviados para a entidade organizadora, que os analisa, compara e emite o relatório individual de desempenho. Um sumário anonimizado dos resultados obtidos pelo grupo de laboratórios participantes é geralmente também disponibilizado.

A participação nos EQA permite ao laboratório a identificação de erros e consequente aplicação de ações corretivas e preventivas, sejam eles ao nível dos métodos, equipamentos ou ao nível de formação técnica [41]. A avaliação do desempenho interlaboratorial, permite ainda verificar a uniformização dos resultados. Dada a grande diversidade de procedimentos e marcas disponibilizadas no mercado, a harmonização dos resultados torna-se um aspeto de elevada relevância.

Apesar dos ensaios EQA serem voluntários, alguns podem ser obrigatórios no processo de acreditação. Da mesma forma, alguns programas podem ser gratuitos, sendo uma boa opção para laboratórios com poucos recursos, muitas vezes disponibilizados por programas nacionais ou internacionais de melhoria da qualidade.

Adicionalmente, a participação nos EQA, tal como a acreditação e certificação, auxilia a visibilidade externa e contribui para a boa imagem dos laboratórios. Ajuda a assegurar junto dos clientes o esforço do laboratório ao nível da qualidade e que este produz resultados fiáveis.

Alguns exemplos de programas na área da Genética incluem os disponibilizados pelo Instituto Nacional de Saúde Dr. Ricardo Jorge (INSA), *European Molecular Quality Network* (EMQN), *Genomics Quality Assessment* (GenQA), *LABQUALITY* e *College of American Pathologists* (CAP). Os laboratórios podem ainda procurar outras formas de certificação, como a disponibilizada pela participação no programa de qualidade laboratorial da *Clinical Laboratory Improvement Amendments* (CLIA).

4. Benchmarking para identificação de variantes

Como foi referido atrás (ponto 9.) existem inúmeras ferramentas e *pipelines* de chamada de variantes genómicos, cada um deles com vantagens e limitações. O *benchmarking* desses recursos torna-se crucial para avaliar o seu desempenho, precisão e confiabilidade.

O *benchmarking* fornece uma estrutura de avaliação padronizada para identificar pontos fortes, fraquezas e vieses associados a diferentes ferramentas e conjuntos de dados. Os recursos de *benchmarking* têm várias finalidades. Em primeiro lugar, ajudam investigadores e clínicos a selecionar ferramentas adequadas para as suas necessidades específicas, considerando fatores como precisão, sensibilidade, especificidade, requisitos computacionais e escalabilidade. Em segundo lugar, o *benchmarking* permite a identificação de limitações e potencial enviesamento nas ferramentas de sequenciação de variantes, permitindo o refinamento de algoritmos e o melhoramento do desempenho em geral. Por fim, o *benchmarking* facilita o estabelecimento de conjuntos de dados de referência e métricas de desempenho que podem ser usados para comparar futuras ferramentas e avanços neste campo. Abaixo, descrevem-se alguns dos recursos de *benchmarking* disponíveis.

A. Conjuntos de dados públicos

O *benchmarking* de ferramentas de sequenciação de variantes frequentemente requer conjuntos de dados curados com variantes conhecidas. Recursos disponíveis publicamente, como o projeto Genome in a Bottle (GIAB) (<https://www.nist.gov/programs-projects/genome-bottle>) [32], fornece um conjunto de dados padronizados para avaliar a precisão e sensibilidade das ferramentas de sequenciação de variantes. Esta base de dados abrange uma ampla variedade de variações genéticas, incluindo variantes de nucleotídeo único (SNVs), inserções, deleções e variantes estruturais. Estas iniciativas facilitam as colaborações entre investigadores para comparar e melhorar o desempenho de diferentes ferramentas. A Aliança Global para Genómica e Saúde estabeleceu uma estrutura de boas práticas

para orientar as avaliações de precisão na chamada de variantes, usando esses recursos [33].

B. Conjuntos de dados sintéticos

Os conjuntos de dados sintéticos, gerados com variantes conhecidas, são recursos valiosos para o *benchmarking* [34]. Permitem que os investigadores controlem vários parâmetros, como profundidade de leitura, frequência alélica de variantes e taxas de erro de sequenciação, e possibilitam uma avaliação detalhada das ferramentas de sequenciação de variantes, em condições controladas, fornecendo ainda uma base para comparações de desempenho.

O *benchmarking* de ferramentas de sequenciação de variantes requer uma consideração cuidadosa de vários fatores, que incluem a escolha de conjuntos de dados, de genomas de referência, e de algoritmos de alinhamento e identificação de variantes, bem como das métricas de desempenho. É crucial garantir que os recursos de *benchmarking* sejam representativos da diversidade genética e da complexidade da população-alvo. Além disso, o *benchmarking* deve abranger múltiplos aspetos, incluindo a eficiência computacional, a escalabilidade e a capacidade de lidar com vários tipos de variantes. À medida que a tecnologia NGS evolui, novos algoritmos e *pipelines* de sequenciação de variantes continuam a emergir. Os recursos de *benchmarking* devem por isso adaptar-se para avaliar o desempenho desses avanços. Esforços colaborativos, conjuntos de dados padronizados e o aprimoramento contínuo das metodologias de *benchmarking* desempenharão um papel vital na precisão e confiabilidade das ferramentas de sequenciação de variantes.

(Página deixada propositadamente em branco)

Bibliografia

- [1] H. L. Rehm *et al.*, 'ACMG clinical laboratory standards for next-generation sequencing', *Genetics in medicine*, vol. 15, no. 9, pp. 733–747, 2013.
- [2] G. Matthijs *et al.*, 'Guidelines for diagnostic next-generation sequencing', *European Journal of Human GWGenetics*, vol. 24, no. 1, pp. 2–5, 2016.
- [3] F. Luh and Y. Yen, 'FDA guidance for next generation sequencing-based testing: balancing regulation and innovation in precision medicine', *NPJ Genom Med*, vol. 3, no. 1, p. 28, 2018.
- [4] S. Hume *et al.*, 'CCMG practice guideline: laboratory guidelines for next-generation sequencing', *J Med Genet*, vol. 56, no. 12, pp. 792–800, 2019.
- [5] M. M. Weiss *et al.*, 'Best Practice Guidelines for the Use of Next-Generation Sequencing Applications in Genome Diagnostics: A National Collaborative Study of Dutch Genome Diagnostic Laboratories', *Hum Mutat*, vol. 34, no. 10, pp. 1313–1321, 2013.
- [6] UNESCO, 'Declaração Universal sobre Bioética e Direitos Humanos'. pp. 1–12, 2006. [Online]. Available: https://unesdoc.unesco.org/notice?id=p::usmarcdef_0000146180_por
- [7] 'Lei n.º 14/90, de 9 de junho', Lisboa: Diário da República n.º 133/1990, Série I de 1990-06-09, pp. 2516–2517.
- [8] UNESCO, 'Declaração Internacional sobre os Dados Genéticos Humanos: recolha, tratamento, utilização, conservação'. pp. 1–17, 2004. [Online]. Available: https://unesdoc.unesco.org/notice?id=p::usmarcdef_0000136112_por
- [9] UNESCO, 'Declaração Universal sobre o Genoma Humano e os Direitos Humanos', 1997. [Online]. Available: <https://unesdoc.unesco.org/ark:/48223/pf0000110220.page=47>
- [10] 'Regulamento (UE) 2016/679', Bruxelas: Parlamento Europeu e do Conselho, de 27 de abril de 2016, pp. 1–88.

- [11] ‘Lei n.º 58/2019, de 8 de agosto’, Lisboa: Diário da República n.º 151/2019, Série I de 2019-08-08, pp. 3–40.
- [12] European Commission, ‘Comunicação da Comissão ao Parlamento Europeu, ao Conselho, ao Comité Económico e Social Europeu e ao Comité das Regiões: Construir o futuro digital da Europa, COM (2020) 67 final’, Bruxelas, 2020. [Online]. Available: <https://eur-lex.europa.eu/legal-content/PT/TXT/?uri=CELEX:52020DC0067>
- [13] ‘Lei n.º 12/2005, de 26 de janeiro’, Lisboa: Diário da República n.º 18/2005, Série I-A de 2005-01-26, pp. 606–611.
- [14] ‘Lei n.º 21/2014, de 16 de abril’, Lisboa: Diário da República n.º 75/2014, Série I de 2014-04-16, pp. 2450–2465.
- [15] ‘Decreto-Lei n.º 80/2018, de 15 de outubro’, Lisboa: Diário da República n.º 198/2018, Série I de 2018-10-15, pp. 4965–4970.
- [16] ‘Despacho n.º 10009/2019, de 5 de novembro’, Lisboa: Diário da República n.º 212/2019, Série II de 2019-11-05, pp. 66–80.
- [17] United Nations, ‘Recommendations on the Transport of Dangerous Goods’, New York & Geneva, 2019.
- [18] World Health Organization, ‘Guidance on regulations for the transport of infectious substances 2021-2022’, 2021.
- [19] World Health Organization, ‘Manual de Biossegurança Laboratorial – 4ª edição’, 2021.
- [20] World Health Organization, ‘World Health Organization’. <https://www.who.int/> (accessed May 21, 2023).
- [21] ‘Decreto-Lei n.º 84/97, de 16 de abril’, Lisboa: Diário da República n.º 89/1997, Série I-A de 1997-04-16, pp. 1702–1709.
- [22] ‘Decreto-Lei n.º 102-A/2020, de 9 de dezembro’, Lisboa: Diário da República n.º 238/2020, 1º Suplemento, Série I de 2020-12-09, pp. 2–50.
- [23] World Health Organization, ‘Manual de Segurança Biológica em Laboratório – 3ª edição’, Genebra, 2004.
- [24] Illumina, ‘Index Adapters: Pooling Guide, Document 1000000041074 v11’, 2021.
- [25] A. M. Bolger, M. Lohse, and B. Usadel, ‘Trimmomatic: a flexible trimmer for Illumina sequence data’, *Bioinformatics*, vol. 30, no. 15, pp. 2114–2120, 2014.
- [26] H. Li and R. Durbin, ‘Fast and accurate long-read alignment with Burrows–Wheeler transform’, *Bioinformatics*, vol. 26, no. 5, pp. 589–595, 2010.

- [27] H. Li *et al.*, ‘The sequence alignment/map format and SAMtools’, *bioinformatics*, vol. 25, no. 16, pp. 2078–2079, 2009.
- [28] M. H.-Y. Fritz, R. Leinonen, G. Cochrane, and E. Birney, ‘Efficient storage of high throughput DNA sequencing data using reference-based compression’, *Genome Res*, vol. 21, no. 5, pp. 734–740, 2011.
- [29] R. Poplin *et al.*, ‘A universal SNP and small-indel variant caller using deep neural networks’, *Nat Biotechnol*, vol. 36, no. 10, pp. 983–987, 2018.
- [30] P. Danecek *et al.*, ‘The variant call format and VCFtools’, *Bioinformatics*, vol. 27, no. 15, pp. 2156–2158, 2011.
- [31] S. Richards *et al.*, ‘Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology’, *Genetics in medicine*, vol. 17, no. 5, pp. 405–423, 2015.
- [32] J. M. Zook *et al.*, ‘Integrating human sequence data sets provides a resource of benchmark SNP and indel genotype calls’, *Nat Biotechnol*, vol. 32, no. 3, pp. 246–251, 2014.
- [33] P. Krusche *et al.*, ‘Best practices for benchmarking germline small-variant calls in human genomes’, *Nat Biotechnol*, vol. 37, no. 5, pp. 555–560, 2019.
- [34] H. Li *et al.*, ‘A synthetic-diploid benchmark for accurate variant-calling evaluation’, *Nat Methods*, vol. 15, no. 8, pp. 595–597, 2018.
- [35] Association of public health laboratories (APHL), ‘Next Generation Sequencing Implementation Guide’, Silver Spring, 2016.
- [36] IPAC, ‘Guia para a aplicação da NP EN ISO 15189’, 2017.
- [37] ‘Portaria n.º 167/2014, de 21 de agosto’, Lisboa: Diário da República n.º 160/2014, Série I de 2014-08-21, pp. 4382–4392.
- [38] ‘Despacho n.º 14303/2022, de 14 de dezembro’, Lisboa: Diário da República n.º 239/2022, Série II de 2022-12-14, pp. 68–68.
- [39] ‘Despacho n.º 14174/2016, de 25 de novembro’, Lisboa: Diário da República n.º 227/2016, Série II de 2016-11-25, pp. 35123–35123.
- [40] ‘Portaria n.º 392/2019, de 5 de novembro’, Lisboa: Diário da República n.º 212/2019, Série I de 2019-11-05.
- [41] World Health Organization, ‘Overview of External Quality Assessment (EQA)’, 2009.

(Página deixada propositadamente em branco)

OBRA PUBLICADA
COM A COORDENAÇÃO
CIENTÍFICA



universidade
de aveiro

1 2



9 0



IMPRESA DA
UNIVERSIDADE
DE COIMBRA
COIMBRA UNIVERSITY PRESS