Review

Next-Generation Sequencing (NGS) and Third-Generation Sequencing (TGS) for the Diagnosis of Thalassemia

Syahzuwan Hassan [1,2], **Rosnah Bahar** [1], **Muhammad Farid Johan** [1], **Ezzeddin Kamil Mohamed Hashim** [3], **Wan Zaidah Abdullah** [1], **Ezalia Esa** [2], **Faidatul Syazlin Abdul Hamid** [2] and **Zefarina Zulkafli** [1,*]

1   Department of Hematology, School of Medical Sciences, Health Campus, Universiti Sains Malaysia, Kubang Kerian 16150, Malaysia
2   Institute for Medical Research, Shah Alam 40170, Malaysia
3   School of Health Sciences, Universiti Sains Malaysia, Kubang Kerian 16150, Malaysia
*   Correspondence: zefarinazulkafli@gmail.com

Abstract: Thalassemia is one of the most heterogeneous diseases, with more than a thousand mutation types recorded worldwide. Molecular diagnosis of thalassemia by conventional PCR-based DNA analysis is time- and resource-consuming owing to the phenotype variability, disease complexity, and molecular diagnostic test limitations. Moreover, genetic counseling must be backed-up by an extensive diagnosis of the thalassemia-causing phenotype and the possible genetic modifiers. Data coming from advanced molecular techniques such as targeted sequencing by next-generation sequencing (NGS) and third-generation sequencing (TGS) are more appropriate and valuable for DNA analysis of thalassemia. While NGS is superior at variant calling to TGS thanks to its lower error rates, the longer reads nature of the TGS permits haplotype-phasing that is superior for variant discovery on the homologous genes and CNV calling. The emergence of many cutting-edge machine learning-based bioinformatics tools has improved the accuracy of variant and CNV calling. Constant improvement of these sequencing and bioinformatics will enable precise thalassemia detections, especially for the CNV and the homologous *HBA* and *HBG* genes. In conclusion, laboratory transiting from conventional DNA analysis to NGS or TGS and following the guidelines towards a single assay will contribute to a better diagnostics approach of thalassemia.

Keywords: PCR; thalassemia; sequencing; NGS; TGS; CNV

1. Introduction

The earliest clinical description of thalassemia was by the Detroit pediatricians Thomas B. Cooley and Pearl Lee, who defined a severe form of anemia among children with splenomegaly and bone deformation [1], which was later termed Cooley's anemia. Whipple and Bradford coined the term "thalassemia" in 1936 when they described Cooley's anemia with erythroblastic anemia [2]. The name originated from the Greek words meaning "sea" and "blood", as the anemic patients were all of Mediterranean origin. Later, the disease was found widely throughout the Indian subcontinent, SEA, and the Middle East.

Thalassemia and hemoglobinopathies are genetic disorders caused by gene defects that hinder the normal production of hemoglobin, the main protein found in RBCs responsible for binding and carrying oxygen-RBCs from the lungs to the various body tissues. Normal adult hemoglobin consists of two pairs of α and β-chains, respectively. Synthesis of these proteins is coordinated to ensure equal production levels in erythropoietic cells [3]. When one or both copies fail to produce normal β-globin, the α-globin gene continues its normal α-globin production. The consequence of these mutations is an imbalance of α/β-globin chain synthesis, evidently in the homozygous forms, leading to the accumulation of free α-globin chains forming highly toxic aggregates [4].

Both alpha (α)- and beta (β)-thalassemia phenotypes are classified separately. α-thalassemia is classified into four phenotypes: the silent carrier, trait, HbH, and Hb Bart's.

Meanwhile, β-thalassemia is classified into four phenotypes: silent carrier (β^{++}), carrier, intermedia, and major. β-thalassemia minor or trait refers to carriers of a β-thalassemia mutation with a normal β-globin on the other allele. They are clinically asymptomatic. β-thalassemia major is characterized by infancy-onset severe anemia that requires lifelong blood transfusion for survival [5]. In β-thalassemia major, the excess unpaired α-globin chains aggregate to form inclusion bodies that damage RBC membranes, causing intravascular hemolysis, damage and apoptosis of erythroid precursors, and ineffective erythropoiesis [6]. β-thalassemia intermedia is defined as having an intermediate thalassemia condition between minor and major [7]. Most patients with β-thalassemia intermedia are homozygotes or compound heterozygotes for β-thalassemia [8]. Rarely, a simple carrier of β-thalassemia can be symptomatic, such as in the case of co-inheritance of segmental duplication of the α-chain, which increases the imbalance ratio of β- and α-tetramer, resulting in a more severe phenotype [9,10].

The separate phenotype classification of α- and β-thalassemia can be perplexing, especially for patients with a complex genotype such as concomitant α- and β-thalassemia. Thus, new symptomatic thalassemia called transfusion-dependent thalassemia (TDT) and non-transfusion-dependent thalassemia (NTDT) were proposed primarily to classify various α- and β-thalassemia and variants [11]. TDT refers to patients requiring lifelong regular blood transfusions to survive [12]. NTDT is a term to describe patients that do not require such lifelong regular transfusions for survival, although they may require occasional or even frequent transfusions in certain clinical settings and for defined periods [13]. NTDT encompasses five clinically distinct forms: β-thalassemia intermedia, hemoglobin E/β-thalassemia (mild and moderate forms), α-thalassemia intermedia (Hb H disease), hemoglobin S/β-thalassemia, and hemoglobin C thalassemia [11].

Mutations in the *HBA* and the *HBB* are associated with α- and β-thalassemia or variants, respectively. They are extremely heterogeneous, with more than a thousand types of mutations comprising single nucleotide variation (SNV), indels, segmental deletions and duplications, and segmental inversion [14–16]. The β-thalassemia is mainly caused by SNV, while segmental deletions are the common cause of α-thalassemia. Deletions of the *HBB* cluster lead to either high persistence of fetal hemoglobin (HPFH), delta beta (δβ-thalassemia), or β-thalassemia and rarely are concomitant with segmental inversions and donor insertion. Segmental duplications are more common in the α-globin cluster than in the β-globin cluster. The importance of molecular diagnosis in thalassemia prevention and its utilization in the clinical decision has become a standard practice in patient management. The variable phenotypic expression renders molecular study crucial for the confirmation of diagnosis, treatment, and counseling purposes.

2. Conventional DNA Analysis

Before the invention of PCR, the diagnosis of β-thalassemia was performed by linkage analysis using restriction fragment length polymorphism (RFLP) [17,18], and Southern transfer and hybridization or Southern-blot analysis [19,20]. Although the enzymatic amplification of DNA was introduced in 1971 [21], the clinical application was described years later for prenatal diagnosis of sickle cell anemia. Using less than 100 times of DNA or as little as 20 ng of DNA, improved sensitivity of DNA hybridization to Phosphorus-32 (^{32}P) end-labeled oligonucleotide probe (isotope) was achieved [22]. The enzymatic amplification of DNA was later termed polymerase chain reaction (PCR) [23]. This enabled the use of a non-radioactive probe of horseradish peroxidase-labeled oligonucleotides for the dot-blot analysis [24,25].

2.1. Reverse Dot-Blot Analysis

Later, the dot-blot analysis was reversed. The horseradish peroxidase-labeled sequence-specific oligonucleotide probes were spotted onto the nylon membrane, allowing simultaneous hybridization reaction of an entire series of sequences [26]. Customized reverse dot blot has been used in many populations [27–34]. Typically, a reverse dot-blot hy-

bridization requires three steps; immobilization of allele-specific oligonucleotide probe to a nylon membrane, amplification of the targeted region of DNA using a biotinylated primer, and hybridization of the biotinylated DNA to the probe-bound nylon membrane using streptavidin-alkaline phosphatase and color substrates.

2.2. Gap-PCR

Gap-PCR is the amplification of the excess segment of deletional thalassemia types. By amplifying the excess segment, an estimation of deletional types is made based on the size of the amplicons. The most commonly used gap-PCR is the multiplex gap-PCR detecting common deletional α-thalassemia [35] and $\alpha\alpha\alpha^{\text{anti 3.7}}$ and $\alpha\alpha\alpha^{\text{anti 4.2}}$ triplications [36]. Other gap-PCR useful to detect deletional forms of α- and β-thalassemia include the gap-PCR for the HPFH, $\delta\beta$-thalassemia, and β-thalassemia [37–39].

2.3. Amplification Refractory Mutation System (ARMS) or Allele-Specific Polymerase Chain Reaction (ASPCR)

Wu and his colleagues proposed a simple allele-specific oligonucleotide PCR (ASPCR) approach that did not require enzyme digestion and blot hybridization for DNA analysis of sickle cell anemia [40]. Later, the same concept dubbed amplification refractory mutation system (ARMS) was described [41]. It was based on allele-specific primers with a modified 3′ end to exactly match the point mutation of choice and the inclusion of a mismatch at the fourth nucleotide of the 3′ end to deliver the extra specificity. The lack of 3′ to 5′ exonuclease activity of Taq polymerase reduced its ability to extend efficiently. These primers can amplify SNV and small indel mutations, whereas homozygosity or heterozygosity is detectable using oligonucleotide primers that perfectly match the wild-type sequences at the same position of the same mutation. Typically, an ARMS requires the amplification of DNA fragments by allele-specific PCR followed by gel electrophoresis of the amplified DNA. Customized multiplex ARMS have also been used in many populations [33,41–52].

2.4. Sanger Sequencing

In 1975, a new technique allowing the determination of sequence in a specific region of the DNA chain was described [53]. Later, the isotope labeling was replaced by fluorophore-labeled dNTPs that were covalently attached to the oligonucleotide primer used in enzymatic DNA sequence analysis [54]. Then, the autoradiography step was replaced by a computer that acquires the sequence information directly [55]. With the introduction of PCR, sequencing of specific regions enabled the detection of rare β-thalassemia [56,57]. In 1990, the separation of fluorescently labeled DNA fragments by capillary electrophoresis (CE) was introduced [58]. Sequencing was used as a subsequent method following targeted PCR, RFLP, and dot-blot analyses [29,33,50,51,59], and became the gold-standard for diagnosis of thalassemia mutations and copy number variation (CNV) breakpoint analysis.

2.5. Multiplex Ligation Probe-Dependent Analysis

Ligation-dependent PCR was introduced for the detection of the Hepatitis C virus [60]. In the same year, a ligation-dependent PCR was patented as multiplex ligations-dependent amplification (MLPA) [61]. Later, MLPA detection of multiple diseases was presented [62] and made commercially available for hereditary disorders, tumor profiling, and methylation status in diagnostic and research by the MRC Holland (Amsterdam, The Netherlands). Each MLPA probe set comprises two probes. The first half contains target-match attached probes with a universal primer. The other half has an extra stuffer sequence meant to produce uniquely sized amplicon upon hybridization, ligation, and amplification. Upon hybridization to the target DNA, ligation of both probes permits measurable amplification of the probes. For SNP detection, the 3′ mutation-specific MLPA probe produced a quantifiable signal when annealed to the target sequence. Using the MLPA probe mix (MRC-Holland, Amsterdam, The Netherlands), α-thalassemia [63,64], large segmental

duplications among symptomatic β-thalassemia carriers [65–69], *HBG1-HBG2* deletion [64], and εγδβ-thalassemia [70,71] were discovered.

3. Advanced Molecular Techniques towards the Single-Assay DNA Analysis

For many years, DNA analyses of Mendelian disease such as thalassemia have relied on two major steps: identifying common mutations via various targeted DNA analysis methods, and rare mutation discovery via Sanger sequencing and MLPA analysis. An extensive DNA analysis detects *HBA*, *HBB*, and *HBD* variants and CNV via Sanger sequencing and MLPA, respectively (Figure 1). Indels are common in *HBB*, thus extra reads are required to accommodate the frameshifted heterozygous mixed Sanger sequencing reads. Homozygous variants must consider the possibility of accompanying deletion on the other allele, so an additional test to rule it out is a must.
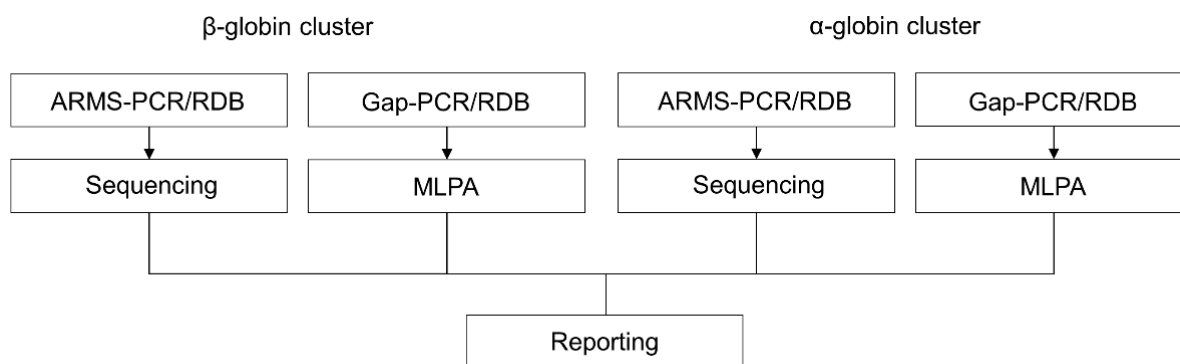

Figure 1. Comprehensive conventional PCR-based genotyping. ARMS-PCR and gap-PCR detect common SNV/indel and CNV, respectively, while the reverse dot-blot (RDB) can be customized to detect SNV/indel and CNV simultaneously. Sequencing and MLPA genotype unknown SNV/indel and CNV, respectively.

For the detection of deletional thalassemia, amplification of the *HBA2* gene [35] and cluster-spanning amplicon were used [37–39] to represent the normality of the other allele, respectively. Homozygous deletional detected by these Gap-PCR must be accompanied by MLPA to rule out rare CNV. MLPA uses a comparative sample to known reference analysis and spanning probe patterns for the identification of CNV. In most cases, the interpretation is straightforward, but cannot be used as a standalone technique because of its limited probes. For deletions and duplications spanning across the *HBA* and *HBB* clusters, validations by long-range PCR are possible because of the many probes available within these clusters. However, larger deletions and duplications breakpoints are harder to estimate owing to the limited number of the MLPA probes available beyond *HBA* and *HBB* clusters.

3.1. Next-Generation Sequencing (NGS)

Two years after the completion of the Human Genome Project (HGP), new sequencing technology emerged. Next-generation sequencing (NGS) enabled scalable high throughput sequencing by adapting parallel detection of small DNA fragments (150–1000 bp). Complete sequencing can be achieved by whole genome sequencing (WGS), while enrichment enables exome, targeted, RNA, and methylation sequencing. NGS chemistry differs between platforms. For instance, Illumina uses clonal array formation and a proprietary reversible terminator for large-scale sequencing. Adaptor and index ligated DNA fragments are immobilized on flowcell, where each fragmented DNA is isothermally amplified, generating millions of DNA clusters. The clusters are excited by light source and the characteristic fluorescent signal is emitted (sequence by synthesis, SBS). The number of cycles determines the length of the reads. The emission wavelength and the signal intensity determine the basecalls (Figure 2).
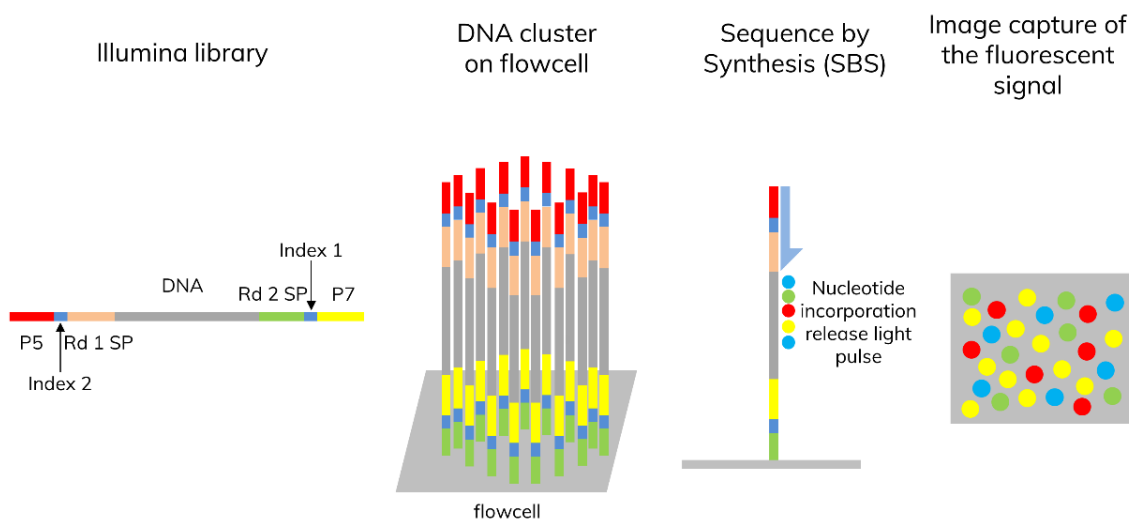
Figure 2. Illumina sequencing by synthesis (SBS). DNA templates are immobilized on a flowcell. When nucleotides are incorporated onto the DNA strands, they release light pulses that are captured by the sequencer and output the bases read from each cluster, along with the quality metrics for each base.

Prior to variant calling, reads are mapped against the reference sequence to identify the region of origin for each sequencing read. Among the most widely used mappers are the Genome Analysis Toolkit (GATK) [72], preferred Burrow-Wheeler Aligner-Maximal exact match (BWA-MEM) [73], Bowtie2 [74], minimap2 [75], and Scalable Nucleotide Alignment Program (SNAP) [76]. SNAP conveniently compresses, sorts, mark-duplicates, and indexes the final output. Various highly accurate variant and indel calling tools employ different approaches and outputs. GATK [72], FreeBayes [77], and SAMtools [78] rely on bayesian approaches. DeepVariant utilized deep neural networks [79], while Strelka2 [80] used a novel mixture-model-based estimation to call variants and indels.

Unlike variants and indel genotyping, detection of large rearrangements in copy-number variants (CNV) from NGS data is challenging owing to the technology's natural limitations such as short read lengths and GC-content bias [81]; however, options are aplenty. Control-FREEC uses a LASSO-based algorithm [82]; DELLY uses short and long-range paired-end mapping and split-read analysis [83]; and CNVkit uses in-target and off-target regions, bias correction using rolling median, and circular binary segmentation (CBS) [84] to call CNV. Tools dedicated to analyzing CNV from exome sequencing reads include ExomeDepth, which uses a beta-binomial model to generate a likelihood value, the hidden Markov model (HMM) to combine the likelihood across multiple exons, and maximum likelihood Viterbi algorithm to provide a set of calls for each sample [85]; CoNIFER uses combined read-depth with singular value decomposition (SVD) normalization and ± 1.5 SVD-transformed standardized z-scores reads per thousand bases per million read sequenced (ZRPKM) (copy number inference from exome reads) [86]; and FishingCNV uses principal component analysis (PCA) of the RPKM, CBS test sample, and comparing segment coverage against the control set distribution [87]. FishingCNV is also available in a graphical software package. The CNV tools and their features are summarized in Table 1.

User-friendly interface genotyping CNV is also available. Detection and Annotation of Copy Number Variations (DeAnnCNV) is an online detection and annotation of CNVs from exome sequencing data that extracts the shared CNVs among multiple samples and provides annotations for the detected CNVs and associated genes [88]. CovCopCan uses normalized read count value (NRC) for each amplicon to generate a CNV detection algorithm based on z-score. False positive is minimized by applying a two-stage ratio to the amplicons, while false negative of two CNV per chromosome is countered by merging the CNV area. A two-dimensional CUSUM chart, local regression curve, and variant call format (VCF) file can be generated [89].

Table 1. CNV tools and their features.

Tool	Algorithm	Highlight
Control-FREEC	LASSO-based, Gaussian mixture models (GMM)	Output BAF from SAM pileup or ratio and copy number calls of each segment Use GC content and mappability profiles to normalize read count if control sample is unavailable
DELLY2	Graph-based paired-end clustering and k-mer filtering for split-read analysis	Call SV from distinct insert sizes PE libraries Output VCF containing SV quality prediction Support short and long reads
CNVkit	Circular Binary Segmentation (CBS), HaarSeg, HMM	Primarily for hybrid capture sequencing Use on- and off-target reads to call CNV Support amplicon sequencing-based TS Multiple segmentation algorithms to choose from
ExomeDepth	Beta-binomial model, HMM, maximum likelihood Viterbi algorithm	An R package works on Windows and UNIX systems Source read count data from multiple samples to build optimized reference sets
CoNIFER	Singular value decomposition and z-scores reads per thousand bases per million read sequenced (SVD-ZRPKM)	Use Matplotlib and Pyplot to generate arbitrary segment of the SVD-ZRPKM data Calculate batch effect biases by concurrently analyzing multiple samples suitable for large sample sets
FishingCNV	PCA	Support CLI and GUI for Windows and UNIX systems Compare coverage depth in test samples and use PCA to remove batch effect

Another hurdle of the CNV analysis is the interpretation of the text format output that lacks genetic or clinical data annotations, so users need to compare all the positions of CNV and use any annotation tools to determine their genetic meaning [90]. Several tools dedicated to interactive and dynamic visualization of CNV have been developed. Scripps Genome Annotation and Distributed Variant Interpretation Server (SG-ADVISER) is a suite consisting of an annotation pipeline and a web server that provides known and predicted information about genetic variants by performing in-depth annotations and functional predictions for variants and CNVs [91]. A web-based application, inCNV, integrates and prioritizes CNV-tool results with user-friendly interfaces and analyzes the importance of called CNVs by generating CNV annotations from Ensembl, Database of Genomic Variants (DGV), ClinVar, and Online Mendelian Inheritance in Man (OMIM) [90]. reconCNV uses delimited result files from most NGS CNV callers to produce an interactive dashboard that can be visualized as Hyper Text Markup Language (HTML) output [92]. Meanwhile, a web server CNVxplorer provides a functional assessment of CNVs in a clinical diagnostic setting by mining a comprehensive set of clinical, genomic, and epigenomic features associated with CNVs [93]. In R, called copy number and beta allele frequency (BAF) data can be visualized by KaryoploteR [94], Gviz [95], and CopyNumberPlots [96].

Targeted sequencing (TS) is the most economical approach for thalassemia as the clusters are small. Ideally, a TS should be able to detect point mutations in *HBA*, *HBB*, *HBD*, and *HBG* and include uniform reads covering the adjacent genes for rare CNV discovery and breakpoint estimation. Adding CNV spanning reads allows accurate direct detection of common CNV. TS has been described to be superior to the conventional DNA analysis strategy. Using the GATK variant calling pipeline coupled with an in-house CNV tool, simultaneous genotyping of globin genes and genetic modifiers had better performance than traditional routine screening [97]. TS of the *HBB* gene and common deletional α-thalassemia of -$\alpha^{3.7}$ and -$\alpha^{4.2}$ using Ion Torrent showed genotype concordance to that of the conventional PCR [98]. A combination of gap-PCR genotyping of the common deletional α-thalassemia and variant genotyping of *HBA* and *HBB* genes by NGS produced higher sensitivity detection than using the MCV, MCH, and HbA2 screening strategy [99]. Higher sensitivity and specificity were achieved with prioritized CNV genotyping in the α-globin cluster along with *HBA* and *HBB* variants [100,101]. New *HBB* cluster deletion dubbed Inv-Del English V $\varepsilon\gamma\delta\beta$-thalassemia (HbVar 2935) of 122.6 kb deletion with 56 kb and 82 bp inversion was characterized using customized targeted panel and RPKM analysis [71]. Zebisch et al., identified a novel variant of $\varepsilon\gamma\delta\beta$-thalassemia using MLPA and comparative genomic hybridization (CGH) that was missed by NGS; however, CNV analysis was not

mentioned [102]. Whole exome sequencing (WES)-based CNV analysis using an exome hidden Markov model (XHMM) [103] identified α-globin cluster duplication in severe β-thalassemia carriers [68].

Short reads mapping of highly homologous regions such as in the *HBA1*, *HBA2*, *HBG1*, and *HBG2* genes is still challenging for NGS. A customized bioinformatics pipeline called NGS4THAL incorporates realignment of the ambiguously mapped reads derived from the hemoglobin gene cluster homologs for variants and indels calling, coupled with multiple tools for CNV discovery to improve genotyping sensitivity and specificity [104]. Using customized long-read WGS of 400 bp per read, patients with rare forms of α- and β-thalassemia were diagnosed [105]. Longer-range reads can also be achieved using link-read sequencing. It utilized multiple bar-coding of the gDNA prior to fragmentations. This allows whole-genome phasing that provides haplotype information valuable for genetic diseases. Successful haplotype-phasing for embryo selection in preimplantation genetic testing has been demonstrated using link-read sequencing for an –SEA carrier partner [106].

3.2. Third-Generation Sequencing (TGS)

Third-generation sequencing (TGS) employs single-molecule sequencing (SMS) that directly sequences individual DNA or RNA strands present in a sample of interest without prior clonal amplification of the DNA [107]. Uninterrupted DNA polymerase incorporates fluorescently labeled deoxyribonucleoside triphosphates (dNTPs), generating continuous DNA synthesis [108]. Pacific Bioscience (PacBio) uses single-molecule real-time (SMRT) isoform sequencing (Iso-Seq). DNA library is created by transforming double-stranded DNA with ligation adapters into circular single-stranded DNA (SMRTbell). Base-calling of the SMRTbell occurs in a chip called an SMRT cell that contains a photonic nanostructure called zero-mode waveguide (ZMW) wells. Immobilized polymerase on the surface of each well initiates DNA replication, producing an interpretable fluorescent pulse (Figure 3) [109]. The Sequel II device outputs Circular Consensus Sequence (CCS) reads. CCS features sequencing of the same molecule multiple times, generating multiple subreads of the SMRTbell library called highly accurate long reads (HiFi reads), thus improving the accuracy of SNV calling [110].
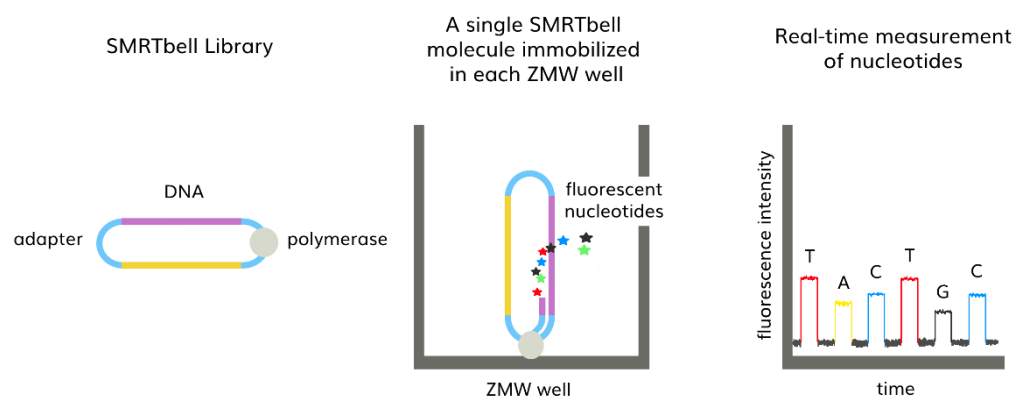

Figure 3. PacBio's SMRTbell library base calling in SMRT cell containing millions of ZMW wells. Anchored polymerase incorporates the nucleotides, emitting light that is measured in real-time. A base-calling algorithm translates the light into DNA sequence.

Meanwhile, in nanopore-based SMS, DNA molecules are individually translocated through nanoscale pores that only permit the passage of single-stranded DNA in a strict linear sequence [107]. Oxford Nanopore Technology (ONT) sequencers measure changes in ionic current when the DNA fragments pass through protein nanopores in a semi-synthetic insulated membrane. A single library comprises a DNA fragment, an adapter-bound motor protein, and a tethering molecule that chains the DNA to the nanopore and membrane [111,112]. Motor protein controls the translocation speed of the DNA, and it feeds DNA bases through the pore. DNA passing through the nanopore leads to a continual

change in current, known as the "squiggle" stored by the MinKNOW ™ software (version 20.10, Oxford, UK). Using a neural network algorithm, MinKNOW translates the squiggle into nucleotides using graphical processing units (GPUs) in real time [113] (Figure 4).
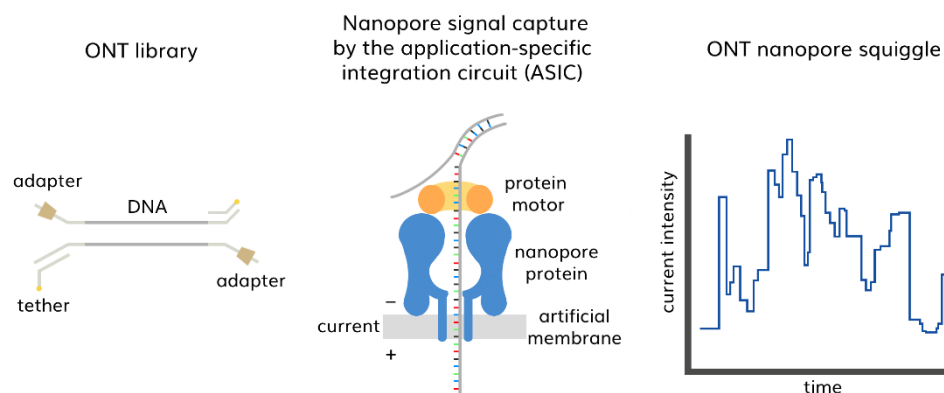

Figure 4. ONT's library base calling in nanopore protein.

Long-read sequencing is superior for CNV detection as a single read can span across exons, genes, pseudogenes, highly duplicated sequences, and CNV, but fell short owing to a higher indel error rate [114,115]. Various state-of-the-art correction tools are available to counter such errors and can be tested during pipeline optimization [116]. Error correction can be done before or after the genome assembly. Assemblers such as Flye [117], wtdbg2 [118], Shasta [119], and CONSENT [120] assemble the raw data using minimap2 pairwise aligner [75] before polishing/correcting the assembly. Conversely, MECAT (in-house aligner) [121], Canu [122] (using MinHash Alignment Process (MHAP) aligner [123]), Falcon [124] (using basic local alignment with successive refinement, BLASR aligner) [125], and NECAT (in-house assembly module) [126] correct error reads and then assemble them.

Improved variant calling using deep neural network (DNN) algorithms for ONT using PEPPER-Margin-DeepVariant [127], NanoCaller [128], and Clair3-trio [129] has been demonstrated. For variant calling of the HiFi reads, accurate variant calling has been shown using GATK Haplotypecaller [72,110], DeepVariant [79], and HELLO [130]. SMRT reads can be analyzed using its in-house PacBio structural variant (SV) calling and analysis tools (pbsv). Numerous CNV callers supporting both SMRT and ONT reads are available. Sniffles2 uses adaptive clustering (repeat aware), followed by a fast consensus sequence, and a coverage-adaptive filter [131], while cuteSV2 implements heuristic signature purification and a specific-designed scanning line [132] to call CNV. Interspersed duplications, tandem duplications, and insertions of novel elements can be detected by SVIM [133], while SV in low read depth WGS can be detected by the neural-network-based algorithm implemented by NanoVar [134]. Generally, both NGS and TGS workflows begin with library preparation, followed by sequencing of the prepared library, quality assessment, and reads' trimming. However, they differ in reads' assembly as TGS requires reads' polishing either before or after mapping, followed by SNV, indels, and SV calling (Figure 5).

Utilizing the CCS, targeted sequencing using specific amplicons aimed at genotyping common -$\alpha^{3.7}$, -$\alpha^{4.2}$, –^{SEA}, and *HBA* and *HBB* SNVs showed a complete concordance to the conventional PCR-based genotyping [135]. Later, the strategy was modified to allow genotyping of more deletional thalassemia forms. Termed a comprehensive analysis of thalassemia alleles (CATSA), the targeted SMRT sequencing was tested on 1759 samples and successfully genotyped common and rare thalassemia SNV, indels, and CNV [136]. Then, CATSA was used to genotype 100 samples, showing abnormal hematological parameters, but were uninformative during conventional genetic diagnosis by RDB and Gap-PCR (genotyping -$\alpha^{3.7}$, -$\alpha^{4.2}$, –^{SEA} only). Ten rare mutations were found [137]. Recently, Li et al., demonstrated detections of HbH disease caused by various deletional and non-deletional α-thalassemia mutations and concomitant α-thalassemia with point mutations and indel types of β-thalassemia [138]. Using MLPA and SMRT, an α-globin gene cluster 27,311 bp

deletion ($-^{27.3}/\alpha\alpha$), an HS-40 region 16,079 bp deletion, a rearrangement of $-\alpha^{3.7}\alpha1\alpha2$ on one allele, a *ß-globin* gene cluster *HBG1-HBG2* 4924 bp deletion, and a 15.8 kb deletion α-thalassemia were characterized [63,64].
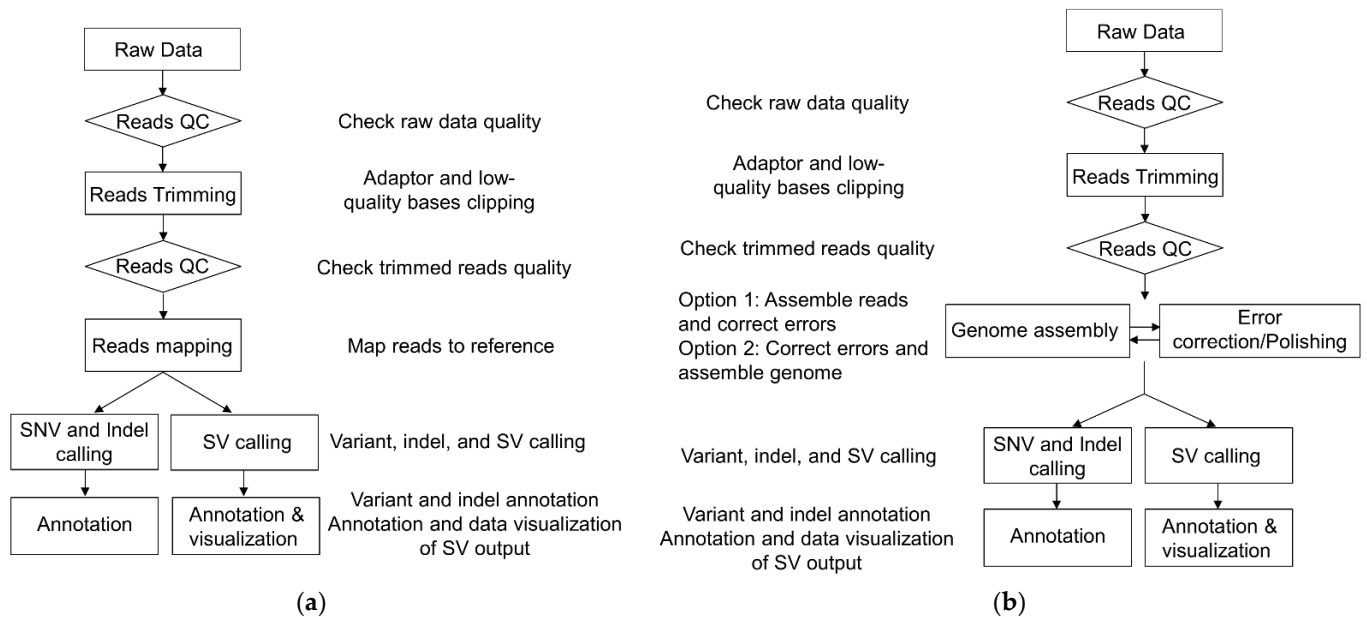

(**a**)　　　　　　　　　　　　　　　　　　　　　(**b**)

Figure 5. Overview of bioinformatics workflow for (**a**) NGS and (**b**) TGS, requiring additional error correction/polishing, either by assembling the genome first followed by error correction or vice versa. Both pipelines use binary alignment map (bam) for SNV, indel, and SV calling. SNV and indel are called simultaneously, while SV calling uses specialized tools/software.

In a non-invasive prenatal testing (NIPT), a long-range 20 kb amplicon sequenced by ONT and NGS was used to phase parental haplotypes to determine fetal inherited haplotypes by the relative haplotype dosage (RHDO) analysis, and successfully genotyped 12 of the 13 fetal thalassemia statuses [139]. Comparative analysis of ONT and Sanger sequencing showed 100% concordance of *HBB* genotyping in a small-scale study in Tanzania [140]. Liu et al., demonstrated genotyping of homozygous—SEA deletion embryos using simple read density plots across the *HBA* locus, thus showing the feasibility of ONT sequencing for preimplantation genetic testing (PGT) [141].

4. Discussion

Conventionally, differential diagnosis examination of hematological parameters and patient's phenotype is used to decide the DNA analysis. It is ambiguous owing to the phenotype variability and the conventional DNA analysis test limitations. Generally, point mutations and indels are detected by ARMS-PCR or sequencing and large deletions are detected by gap-PCR or MLPA. A homozygous β-thalassemia detected by the ARMS-PCR and sequencing may not be a true homozygote, but rather a compound heterozygous with a deletional β-thalassemia or δβ-thalassemia that must be ruled out by the gap-PCR, MLPA, or cascade screening. Misdiagnosis could occur in complex genotyping, which alters the hematological parameters, such as in the mild β-thalassemia/δ-thalassemia with normal HbA$_2$. The multiple methods are labor-intensive and prolong the laboratory turnaround time (TAT). Conversely, NGS and TGS permit simultaneous mutation detections of SNV, indel, and CNV, thus genotyping concomitant α- and β-thalassemia and rare variants. It improves the DNA analysis precision and promotes a better understanding of the genotype–phenotype relationship.

NGS and TGS also allow minimal DNA usage and increase throughput by sample multiplexing, hence reducing per-sample cost and TAT. Substantial costs for NGS and TGS are from the library preparation and sequencing with small amounts by the analysis. For

conventional DNA analysis, the cost comes mainly from the reagents (i.e., PCR master mix) and may be lower than that of NGS and TGS. However, it can never match the resolutions of NGS and TGS. The specific challenge of NGS and TGS is the technically demanding bioinformatics analyses. While TGS permits phasing during genome assembly, haplotype-phasing is possible for short reads using pangenomic mapping. A pangenome integrates whole-genome sequences from multiple individuals to represent genetic diversity [142,143]. The reference pangenome would potentially address the biases and errors of the single linear reference (GRCh38) and is managed by the Human Pangenome Reference Consortium (HPRC) [144]. However, tools and pipelines for graph-based pangenome mapping are more complex and limited. Table 2 outlines the advantages of NGS and TGS over conventional DNA analysis.

Table 2. Comparison of DNA analysis of thalassemia using the conventional method, NGS, and TGS.

Feature	Conventional	NGS	TGS
DNA usage	High	Low	Low
Mutation detection	Method-dependent	Simultaneous	Simultaneous
Haplotype-phasing	Not relevant	Yes [1]	Yes
TAT	Long	Short	Short
Per sample cost	Variable	Uniform	Uniform
Technical difficulty	Low	High	High

[1] Via pangenome mapping.

5. Conclusions

NGS and TGS are useful for simultaneous SNV, indels, and SV genotyping. Owing to the mutation heterogeneities, a single-assay TS requires uniform reads, long reads for haplotype-phasing of the homologous genes (*HBA* and *HBG*), and breakpoint spanning reads for direct detection of the common deletions and duplications. The improvement of the sequencing technology to reduce the error rates and limitations, such as the upcoming Pacbio Revio long read and Onso short read systems, the Illumina Complete Long-Read technology, and state-of-the-art reads error correction tools for TGS, can be leveraged for the detection of the heterogenous thalassemia mutations.

These technologies will not replace conventional screening and PCR-based genotyping thanks to their ease of use. Complex thalassemia such as in the HK$\alpha\alpha$ (Hong Kong $\alpha\alpha$) allele containing both the -$\alpha^{3.7}$ and $\alpha\alpha\alpha^{\text{anti4.2}}$ can be difficult to diagnose by NGS. Furthermore, conventional PCR can validate the genotyping by NGS and TGS. Traditional differential diagnosis by analyzing blood test data before DNA analysis allows genotype–phenotype correlation, thus spotting any human error during sample handling.

The key benefit of NGS and TGS over conventional DNA analysis is in their ability to genotype α- and β-thalassemia simultaneously, allowing complete diagnosis of the thalassemia and the genetic modifiers, which is crucial for genetic counseling. Besides, sequencing data banking permits re-analysis when necessary. Instruments such as the computational infrastructure for the data analysis, a skilled bioinformatics technician, properly documented TS, and the bioinformatics pipeline development and optimization that follow recommended guidelines are expensive for any start-up laboratory, but will ease as sample throughput increases. Towards clinically relevant usage of these technologies, several recommendations and guidelines have been introduced [145–148].

Author Contributions: Conceptualization, S.H.; software, S.H.; writing—original draft preparation, S.H. and F.S.A.H.; writing—review and editing, E.E., Z.Z., M.F.J., W.Z.A., E.K.M.H. and R.B. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

References

1. Cooley, T.B.; Witwer, E.R.; Lee, P. Anemia in Children with Splenomegaly and Peculiar Changes in the Bones Report of Cases. *Am. J. Dis. Child.* **1927**, *34*, 347–363. [CrossRef]

2. Whipple, G.H.; Bradford, W.L. Mediterranean Disease-Thalassemia (Erythroblastic Anemia of Cooley): Associated Pigment Abnormalities Simulating Hemochromatosis. *J. Pediatr.* **1936**, *9*, 279–311. [CrossRef]

3. Ribeil, J.-A.; Arlet, J.-B.; Dussiot, M.; Cruz Moura, I.; Courtois, G.; Hermine, O. Ineffective Erythropoiesis in β-Thalassemia. *Sci. World J.* **2013**, *2013*, 394295. [CrossRef] [PubMed]

4. Khandros, E.; Thom, C.S.; D'Souza, J.; Weiss, M.J. Integrated Protein Quality-Control Pathways Regulate Free α-Globin in Murine β-Thalassemia. *Blood* **2012**, *119*, 5265–5275. [CrossRef]

5. Galanello, R.; Origa, R. Beta-Thalassemia. *Orphanet J. Rare Dis.* **2010**, *5*, 11. [CrossRef]

6. Yuan, J.; Angelucci, E.; Lucarelli, G.; Aljurf, M.; Snyder, L.M.; Kiefer, C.R.; Ma, L.; Schrier, S.L. Accelerated Programmed Cell Death (Apoptosis) in Erythroid Precursors of Patients with Severe β-Thalassemia (Cooley's Anemia). *Blood* **1993**, *82*, 374–377. [CrossRef]

7. Nathan, D.G.; Oski, F.A. *Hematology of Infancy and Childhood*, 4th ed.; W.B. Saunders: Philadelphia, PA, USA, 1993.

8. Galanello, R.; Cao, A. Relationship between Genotype and Phenotype. *Ann. N. Y. Acad. Sci.* **1988**, *850*, 325–333. [CrossRef]

9. Goossens, M.; Dozy, A.M.; Emburyt, S.H.; Zachariadest, Z.; Hadjiminast, M.G.; Stamatoyannopoulos, G.; Kan, Y.W.A.I. Triplicated A-Globin Loci. *Proc. Natl. Acad. Sci. USA* **1980**, *77*, 518–521. [CrossRef]

10. Henni, T.; Belhani, M.; Morle, F.; Bachir, D.; Tabone, P.; Colonna, P.; Godet, J. Alpha Globin Gene Triplication in Severe Heterozygous Beta Thalassemia. *Acta Haematol.* **1985**, *74*, 236–239. [CrossRef]

11. Weatherall, D.J. The Definition and Epidemiology of Non-Transfusion-Dependent Thalassemia. *Blood Rev.* **2012**, *26*, S3–S6. [CrossRef]

12. Rachmilewitz, E.A.; Giardina, P.J. How I Treat Thalassemia. *Blood* **2011**, *118*, 3479–3488. [CrossRef]

13. Musallam, K.M.; Rivella, S.; Vichinsky, E.; Rachmilewitz, E.A. Non-Transfusion-Dependent Thalassemias. *Haematologica* **2013**, *98*, 833–844. [CrossRef] [PubMed]

14. Lederer, C.W.; Basak, A.N.; Aydinok, Y.; Christou, S.; El-Beshlawy, A.; Eleftheriou, A.; Fattoum, S.; Felice, A.E.; Fibach, E.; Galanello, R.; et al. An Electronic Infrastructure for Research and Treatment of the Thalassemias and Other Hemoglobinopathies: The Euro-Mediterranean ITHANET Project. *Hemoglobin* **2009**, *33*, 163–176. [CrossRef] [PubMed]

15. Fokkema, I.F.A.C.; Taschner, P.E.M.; Schaafsma, G.C.P.; Celli, J.; Laros, J.F.J.; den Dunnen, J.T. LOVD v.2.0: The next Generation in Gene Variant Databases. *Hum. Mutat.* **2011**, *32*, 557–563. [CrossRef] [PubMed]

16. Giardine, B.M.; Joly, P.; Pissard, S.; Wajcman, H.; Chui, D.H.K.; Hardison, R.C.; Patrinos, G.P. Clinically Relevant Updates of the HbVar Database of Human Hemoglobin Variants and Thalassemia Mutations. *Nucleic Acids Res.* **2021**, *49*, D1192–D1196. [CrossRef]

17. Kan, Y.W.; Lee, K.Y.; Furbetta, M.; Angius, A.; Cao, A. Polymorphism of DNA Sequence in the β-Globin Gene Region. *N. Engl. J. Med.* **1980**, *302*, 185–188. [CrossRef]

18. Kazazian, H.H.; Phillips, J.A.; Boehm, C.D.; Vik, T.A.; Mahoney, M.J.; Ritchey, A.K. Prenatal Diagnosis of β-Thalassemias by Amniocentesis: Linkage Analysis Using Multiple Polymorphic Restriction Endonuclease Sites. *Blood* **1980**, *56*, 926–930. [CrossRef]

19. Geever, R.F.; Wilson, L.B.; Nallaseth, F.S.; Milner, P.F.; Bittner, M. Direct Identification of Sickle Cell Anemia by Blot Hybridization. *Proc. Natl. Acad. Sci. USA* **1981**, *78*, 5081–5085. [CrossRef]

20. Pirastu, M.; Kan, Y.W.; Cao, A.; Conner, B.J.; Teplitz, R.L.; Wallace, R.B. Prenatal Diagnosis of β-Thalassemia. *N. Engl. J. Med.* **1983**, *309*, 284–287. [CrossRef]

21. Kleppe, K.; Ohtsuka, E.; Kleppe, R.; Molineux, I.; Khorana, H.G. Studies on Polynucleotides. XCVI. Repair Replication of Short Synthetic DNA's as Catalyzed by DNA Polymerases. *J. Mol. Biol.* **1971**, *56*, 341–361. [CrossRef]

22. Saiki, R.K.; Scharf, S.; Faloona, F.; Mullis, K.B.; Horn, G.T.; Erlich, H.A.; Arnheim, N. Enzymatic Amplification of β-Globin Genomic Sequences and Restriction Site Analysis for Diagnosis of Sickle Cell Anemia. *Science* **1985**, *230*, 1350–1354. [CrossRef] [PubMed]

23. Mullis, K.B.; Faloona, F.A.B.T.-M. Specific Synthesis of DNA in Vitro via a Polymerase-Catalyzed Chain Reaction. In *Recombinant DNA Part F*; Academic Press: Cambridge, MA, USA, 1987; Volume 155, pp. 335–350, ISBN 0076-6879.

24. Amselem, S.; Nunes, V.; Vidaud, M.; Estivill, X.; Wong, C.; d'Auriol, L.; Vidaud, D.; Galibert, F.; Baiget, M.; Goossens, M. Determination of the Spectrum of β-Thalassemia Genes in Spain by Use of Dot-Blot Analysis of Amplified β-Globin DNA. *Am. J. Hum. Genet.* **1988**, *43*, 95–100. [PubMed]

25. Saiki, R.K.; Chang, C.-A.; Levenson, C.H.; Warren, T.C.; Boehm, C.D.; Kazazian, H.H.; Erlich, H.A. Diagnosis of Sickle Cell Anemia and β-Thalassemia with Enzymatically Amplified DNA and Nonradioactive Allele-Specific Oligonucleotide Probes. *N. Engl. J. Med.* **1988**, *319*, 537–541. [CrossRef] [PubMed]

26. Saiki, R.K.; Walsh, P.S.; Levenson, C.H.; Erlich, H.A. Genetic Analysis of Amplified DNA with Immobilized Sequence-Specific Oligonucleotide Probes. *Proc. Natl. Acad. Sci. USA* **1989**, *86*, 6230–6234. [CrossRef]

27. Maggio, A.; Giambona, A.; Cai, S.P.; Wall, J.; Kan, Y.W.; Chehab, F.F. Rapid and Simultaneous Typing of Hemoglobin S, Hemoglobin C, and Seven Mediterranean Beta-Thalassemia Mutations by Covalent Reverse Dot-Blot Analysis: Application to Prenatal Diagnosis in Sicily. *Blood* **1993**, *81*, 239–242. [CrossRef]

28. Cai, S.-P.; Wall, J.; Kan, Y.W.; Chehab, F.F. Reverse Dot Blot Probes for the Screening of β-Thalassernia Mutationsin Asians and American Blacks. *Hum. Mutat.* **1994**, *3*, 59–63. [CrossRef]

29. Giambona, A.; Lo Gioco, P.; Marino, M.; Abate, I.; Di Marzo, R.; Renda, M.; Di Trapani, F.; Messana, F.; Siciliano, S.; Rigano, P. The Great Heterogeneity of Thalassemia Molecular Defects in Sicily. *Hum. Genet.* **1995**, *95*, 526–530. [CrossRef]

30. Sutcharitchan, P.; Saiki, R.; Fucharoen, S.; Winichagoon, P.; Erlich, H.; Embury, S.H. Reverse Dot-Blot Detection of Thai β-Thalassaemia Mutations. *Br. J. Haematol.* **1995**, *90*, 809–816. [CrossRef]

31. Sutcharitchan, P.; Saiki, R.; Huisman, T.H.; Kutlar, A.; McKie, V.; Erlich, H.; Embury, S.H. Reverse Dot-Blot Detection of the African-American Beta-Thalassemia Mutations. *Blood* **1995**, *86*, 1580–1585. [CrossRef]

32. Chan, V.; Yam, I.; Chen, F.E.; Chan, T.K. A Reverse Dot-Blot Method for Rapid Detection of Non-Deletion α Thalassaemia. *Br. J. Haematol.* **1999**, *104*, 513–515. [CrossRef]

33. Bashyam, M.D.; Bashyam, L.; Savithri, G.R.; Gopikrishna, M.; Sangal, V.; Devi, A.R.R. Molecular Genetic Analyses of β-Thalassemia in South India Reveals Rare Mutations in the β-Globin Gene. *J. Hum. Genet.* **2004**, *49*, 408–413. [CrossRef] [PubMed]

34. Lin, M.; Zhu, J.-J.; Wang, Q.; Xie, L.-X.; Lu, M.; Wang, J.-L.; Wang, C.-F.; Zhong, T.-Y.; Zheng, L.; Pan, M.-C.; et al. Development and Evaluation of a Reverse Dot Blot Assay for the Simultaneous Detection of Common Alpha and Beta Thalassemia in Chinese. *Blood Cells Mol. Dis.* **2012**, *48*, 86–90. [CrossRef] [PubMed]

35. Chong, S.S.; Boehm, C.D.; Higgs, D.R.; Cutting, G.R. Single-Tube Multiplex-PCR Screen for Common Deletional Determinants of α-Thalassemia. *Blood* **2000**, *95*, 360–362. [CrossRef] [PubMed]

36. Wang, W.; Ma, E.S.; Chan, A.Y.; Prior, J.; Erber, W.N.; Chan, L.C.; Chui, D.H.; Chong, S.S. Single-Tube Multiplex-PCR Screen for Anti-3.7 and Anti-4.2 α-Globin Gene Triplications. *Clin. Chem.* **2003**, *49*, 1679–1682. [CrossRef]

37. Craig, J.; Barnetson, R.; Prior, J.; Raven, J.; Thein, S. Rapid Detection of Deletions Causing Delta Beta Thalassemia and Hereditary Persistence of Fetal Hemoglobin by Enzymatic Amplification. *Blood* **1994**, *83*, 1673–1682. [CrossRef]

38. Nussenzveig, R.H.; Vanhille, D.L.; Hussey, N.; Scott Reading, D.; Agarwal, A.M. Development of a Rapid Multiplex PCR Assay for Identification of the Three Common Hemoglobin-Lepore Variants (Boston-Washington, Baltimore, and Hollandia) and Identification of a New Lepore Variant. *Am. J. Hematol.* **2012**, *87*, 74–75. [CrossRef]

39. Tritipsombut, J.; Phylipsen, M.; Viprakasit, V.; Chalaow, N.; Sanchaisuriya, K.; Giordano, P.C.; Fucharoen, S.; Harteveld, C.L. A Single-Tube Multiplex Gap-Polymerase Chain Reaction for the Detection of Eight β-Globin Gene Cluster Deletions Common in Southeast Asia. *Hemoglobin* **2012**, *36*, 571–580. [CrossRef]

40. Wu, D.Y.; Ugozzoli, L.; Pal, B.K.; Wallace, R.B. Allele-Specific Enzymatic Amplification of Beta-Globin Genomic DNA for Diagnosis of Sickle Cell Anemia. *Proc. Natl. Acad. Sci. USA* **1989**, *86*, 2757–2760. [CrossRef]

41. Newton, C.R.; Graham, A.; Heptinstall, L.E.; Powell, S.J.; Summers, C.; Kalsheker, N.; Smith, J.C.; Markham, A.F. Analysis of Any Point Mutation in DNA. The Amplification Refractory Mutation System (ARMS). *Nucleic Acids Res.* **1989**, *17*, 2503–2516. [CrossRef]

42. Fortina, P.; Dotti, G.; Conant, R.; Monokian, G.; Parrella, T.; Hitchcock, W.; Rappaport, E.; Schwartz, E.; Surrey, S. Detection of the Most Common Mutations Causing Beta-Thalassemia in Mediterraneans Using a Multiplex Amplification Refractory Mutation System (MARMS). *PCR Methods Appl.* **1992**, *2*, 163–166. [CrossRef]

43. Mirasena, S.; Shimbhu, D.; Sanguansermsri, M.; Sanguansermsri, T. Detection of β-Thalassemia Mutations Using a Multiplex Amplification Refractory Mutation System Assay. *Hemoglobin* **2008**, *32*, 403–409. [CrossRef] [PubMed]

44. Hassan, S.; Ahmad, R.; Zakaria, Z.; Zulkafli, Z.; Abdullah, W.Z. Detection of β-Globin Gene Mutations among β-Thalassaemia Carriers and Patients in Malaysia: Application of Multiplex Amplification Refractory Mutation System-Polymerase Chain Reaction. *Malaysian J. Med. Sci.* **2013**, *20*, 13–20.

45. Tan, J.A.M.A.; Tay, J.S.H.; Lin, L.I.; Kham, S.K.Y.; Chia, J.N.; Chin, T.M.; Adb Aziz, N.B.; Wong, H.B. The Amplification Refractory Mutation System (ARMS): A Rapid and Direct Prenatal Diagnostic Technique for β-Thalassaemia in Singapore. *Prenat. Diagn.* **1994**, *14*, 1077–1082. [CrossRef]

46. Chang, J.G.; Liu, H.J.; Huang, J.M.; Yang, T.Y.; Chang, C.P. Multiplex Mutagenically Separated PCR: Diagnosis of Beta-Thalassemia and Hemoglobin Variants. *Biotechniques* **1997**, *22*, 520–527. [CrossRef]

47. Khateeb, B.; Moatter, T.; Shaghil, A.M.; Haroon, S.; Kakepoto, G.N. Genetic Diversity of Beta-Thalassemia Mutations in Pakistani Population. *J. Pakistan Med. Assoc.* **2000**, *50*, 293–296.

48. Eng, B.; Patterson, M.; Walker, L.; Chui, D.H.K.; Waye, J.S. Detection of Severe Nondeletional α-Thalassemia Mutations Using a Single-Tube Multiplex ARMS Assay. *Genet. Test.* **2001**, *5*, 327–329. [CrossRef]

49. Bhardwaj, U.; Zhang, Y.-H.; Lorey, F.; McCabe, L.L.; McCabe, E.R.B. Molecular Genetic Confirmatory Testing from Newborn Screening Samples for the Common African-American, Asian Indian, Southeast Asian, and Chinese β-Thalassemia Mutations. *Am. J. Hematol.* **2005**, *78*, 249–255. [CrossRef]

50. Darwish, H.M.; El-Khatib, F.F.; Ayesh, S. Spectrum of β-Globin Gene Mutations Among Thalassemia Patients in the West Bank Region of Palestine. *Hemoglobin* **2005**, *29*, 119–132. [CrossRef]

51. El-Gawhary, S.; El-Shafie, S.; Niazi, M.; Aziz, M.; El-Beshlawy, A. Study of β-Thalassemia Mutations Using the Polymerase Chain Reaction-Amplification Refractory Mutation System and Direct DNA Sequencing Techniques in a Group of Egyptian Thalassemia Patients. *Hemoglobin* **2007**, *31*, 63–69. [CrossRef]

52. Lacerra, G.; Musollino, G.; Di Noce, F.; Prezioso, R. Clementina Carestia Genotyping for Known Mediterranean α-Thalassemia Point Mutations Using a Multiplex Amplification Refractory Mutation System. *Haematologica* **2007**, *92*, 254–255. [CrossRef]

53. Sanger, F.; Coulson, A.R. A Rapid Method for Determining Sequences in DNA by Primed Synthesis with DNA Polymerase. *J. Mol. Biol.* **1975**, *94*, 441–448. [CrossRef] [PubMed]

54. Smith, L.M.; Fung, S.; Hunkapiller, M.W.; Hunkapiller, T.J.; Hood, L.E. The Synthesis of Oligonucleotides Containing an Aliphatic Amino Group at the 5′ Terminus: Synthesis of Fluorescent DNA Primers for Use in DNA Sequence Analysis. *Nucleic Acids Res.* **1985**, *13*, 2399–2412. [CrossRef] [PubMed]

55. Smith, L.M.; Sanders, J.Z.; Kaiser, R.J.; Hughes, P.; Dodd, C.; Connell, C.R.; Heiner, C.; Kent, S.B.; Hood, L.E. Fluorescence Detection in Automated DNA Sequence Analysis. *Nature* **1986**, *321*, 674–679. [CrossRef] [PubMed]

56. Wong, C.; Dowling, C.E.; Saiki, R.K.; Higuchi, R.G.; Erlich, H.A.; Kazazian, H.H.J. Characterization of Beta-Thalassaemia Mutations Using Direct Genomic Sequencing of Amplified Single Copy DNA. *Nature* **1987**, *330*, 384–386. [CrossRef] [PubMed]

57. Di Marzo, R.; Dowling, C.E.; Wong, C.; Maggio, A.; Kazazian, H.H. The Spectrum of β-Thalassaemia Mutations in Sicily. *Br. J. Haematol.* **1988**, *69*, 393–397. [CrossRef] [PubMed]

58. Drossman, H.; Luckey, J.A.; Kostichka, A.J.; D'Cunha, J.; Smith, L.M. High-Speed Separations of DNA Sequencing Reactions by Capillary Electrophoresis. *Anal. Chem.* **1990**, *62*, 900–903. [CrossRef] [PubMed]

59. Aulehla-Scholz, C.; Basaran, S.; Agaoglu, L.; Arcasoy, A.; Holzgreve, W.; Miny, P.; Ridolfi, F.; Horst, J. Molecular Basis of Beta-Thalassemia in Turkey: Detection of Rare Mutations by Direct Sequencing. *Hum. Genet.* **1990**, *84*, 195–197. [CrossRef]

60. Hsuih, T.C.; Park, Y.N.; Zaretsky, C.; Wu, F.; Tyagi, S.; Kramer, F.R.; Sperling, R.; Zhang, D.Y. Novel, Ligation-Dependent PCR Assay for Detection of Hepatitis C in Serum. *J. Clin. Microbiol.* **1996**, *34*, 501–507. [CrossRef]

61. Carrino, J.J. Multiplex Ligations-Dependent Amplification. 1996. Available online: http://patent.google.com/patent/WO19960 15271A1/en (accessed on 25 October 2022).

62. Schouten, J.P.; McElgunn, C.J.; Waaijer, R.; Zwijnenburg, D.; Diepvens, F.; Pals, G. Relative Quantification of 40 Nucleic Acid Sequences by Multiplex Ligation-Dependent Probe Amplification. *Nucleic Acids Res.* **2002**, *30*, e57. [CrossRef]

63. Zhong, Z.; Zhong, G.; Guan, Z.; Chen, D.; Wu, Z.; Yang, K.; Chen, D.; Liu, Y.; Xu, R.; Chen, J. A Novel 15.8 Kb Deletion α-Thalassemia Confirmed by Long-Read Single-Molecule Real-Time Sequencing: Hematological Phenotypes and Molecular Characterization. *Clin. Biochem.* **2022**, *108*, 46–49. [CrossRef]

64. Luo, S.; Chen, X.; Zeng, D.; Tang, N.; Yuan, D.; Liu, B.; Chen, L.; Zhong, Q.; Li, J.; Liu, Y.; et al. Detection of Four Rare Thalassemia Variants Using Single-Molecule Realtime Sequencing. *Front. Genet.* **2022**, *13*, 974999. [CrossRef] [PubMed]

65. Harteveld, C.L.; Refaldi, C.; Cassinerio, E.; Cappellini, M.D.; Giordano, P.C. Segmental Duplications Involving the α-Globin Gene Cluster Are Causing β-Thalassemia Intermedia Phenotypes in β-Thalassemia Heterozygous Patients. *Blood Cells Mol. Dis.* **2008**, *40*, 312–316. [CrossRef] [PubMed]

66. Sollaino, M.C.; Paglietti, M.E.; Perseu, L.; Giagu, N.; Loi, D.; Galanello, R. Association of a Globin Gene Quadruplication and Heterozygous β Thalassemia in Patients with Thalassemia Intermedia. *Haematologica* **2009**, *94*, 1445–1448. [CrossRef]

67. Jiang, H.; Liu, S.; Zhang, Y.L.; Wan, J.H.; Li, R.; Li, D.Z. Association of an α-Globin Gene Cluster Duplication and Heterozygous β-Thalassemia in a Patient with a Severe Thalassemia Syndrome. *Hemoglobin* **2015**, *39*, 102–106. [CrossRef]

68. Steinberg-Shemer, O.; Ulirsch, J.C.; Noy-Lotan, S.; Krasnov, T.; Attias, D.; Dgany, O.; Laor, R.; Sankaran, V.G.; Tamary, H. Whole-Exome Sequencing Identifies an α-Globin Cluster Triplication Resulting in Increased Clinical Severity of β-Thalassemia. *Cold Spring Harb. Mol. Case Stud.* **2017**, *3*, a001941. [CrossRef]

69. Clark, B.; Shooter, C.; Smith, F.; Brawand, D.; Steedman, L.; Oakley, M.; Rushton, P.; Rooks, H.; Wang, X.; Drousiotou, A.; et al. Beta Thalassaemia Intermedia Due to Co-Inheritance of Three Unique Alpha Globin Cluster Duplications Characterised by next Generation Sequencing Analysis. *Br. J. Haematol.* **2018**, *180*, 160–164. [CrossRef] [PubMed]

70. Kubiczkova Besse, L.; Sedlarikova, L.; Kryukov, F.; Nekvindova, J.; Radova, L.; Almasi, M.; Pelcova, J.; Minarik, J.; Pika, T.; Pikalova, Z.; et al. Combination of serum microRNA-320a and microRNA-320b as a marker for Waldenström macroglobulinemia. *Am. J. Hematol.* **2015**, *90*, E51–E52. [CrossRef] [PubMed]

71. Shooter, C.; Rooks, H.; Thein, S.L.; Clark, B. Next Generation Sequencing Identifies a Novel Rearrangement in the HBB Cluster Permitting To-the-Base Characterization. *Hum. Mutat.* **2015**, *36*, 142–150. [CrossRef]

72. DePristo, M.A.; Banks, E.; Poplin, R.; Garimella, K.V.; Maguire, J.R.; Hartl, C.; Philippakis, A.A.; del Angel, G.; Rivas, M.A.; Hanna, M.; et al. A Framework for Variation Discovery and Genotyping Using Next-Generation DNA Sequencing Data. *Nat. Genet.* **2011**, *43*, 491–498. [CrossRef]

73. Li, H. Aligning Sequence Reads, Clone Sequences and Assembly Contigs with BWA-MEM. *arXiv* **2013**, arXiv:1303.3997v2. [CrossRef]

74. Langmead, B.; Salzberg, S.L. Fast Gapped-Read Alignment with Bowtie 2. *Nat. Methods* **2012**, *9*, 357–359. [CrossRef] [PubMed]

75. Li, H. Minimap2: Pairwise Alignment for Nucleotide Sequences. *Bioinformatics* **2018**, *34*, 3094–3100. [CrossRef] [PubMed]

76. Zaharia, M.; Bolosky, W.J.; Curtis, K.; Fox, A.; Patterson, D.A.; Shenker, S.; Stoica, I.; Karp, R.M.; Sittler, T. Faster and More Accurate Sequence Alignment with {SNAP}. *arXiv* **2011**, arXiv:1111.5572.

77. Garrison, E.; Marth, G. Haplotype-Based Variant Detection from Short-Read Sequencing. *arXiv* **2012**, arXiv:1207.3907.

78. Li, H.; Handsaker, B.; Wysoker, A.; Fennell, T.; Ruan, J.; Homer, N.; Marth, G.; Abecasis, G.; Durbin, R. The Sequence Alignment/Map Format and SAMtools. *Bioinformatics* **2009**, *25*, 2078–2079. [CrossRef]

79. Poplin, R.; Chang, P.-C.; Alexander, D.; Schwartz, S.; Colthurst, T.; Ku, A.; Newburger, D.; Dijamco, J.; Nguyen, N.; Afshar, P.T.; et al. A Universal SNP and Small-Indel Variant Caller Using Deep Neural Networks. *Nat. Biotechnol.* **2018**, *36*, 983–987. [CrossRef]

80. Kim, S.; Scheffler, K.; Halpern, A.L.; Bekritsky, M.A.; Noh, E.; Källberg, M.; Chen, X.; Kim, Y.; Beyter, D.; Krusche, P.; et al. Strelka2: Fast and Accurate Calling of Germline and Somatic Variants. *Nat. Methods* **2018**, *15*, 591–594. [CrossRef]

81. Teo, S.M.; Pawitan, Y.; Ku, C.S.; Chia, K.S.; Salim, A. Statistical Challenges Associated with Detecting Copy Number Variations with Next-Generation Sequencing. *Bioinformatics* **2012**, *28*, 2711–2718. [CrossRef]

82. Boeva, V.; Popova, T.; Bleakley, K.; Chiche, P.; Cappo, J.; Schleiermacher, G.; Janoueix-Lerosey, I.; Delattre, O.; Barillot, E. Control-FREEC: A Tool for Assessing Copy Number and Allelic Content Using next-Generation Sequencing Data. *Bioinformatics* **2012**, *28*, 423–425. [CrossRef]

83. Rausch, T.; Zichner, T.; Schlattl, A.; Stütz, A.M.; Benes, V.; Korbel, J.O. DELLY: Structural Variant Discovery by Integrated Paired-End and Split-Read Analysis. *Bioinformatics* **2012**, *28*, i333–i339. [CrossRef]

84. Talevich, E.; Shain, A.H.; Botton, T.; Bastian, B.C. CNVkit: Genome-Wide Copy Number Detection and Visualization from Targeted DNA Sequencing. *PLoS Comput. Biol.* **2016**, *12*, e1004873. [CrossRef] [PubMed]

85. Plagnol, V.; Curtis, J.; Epstein, M.; Mok, K.Y.; Stebbings, E.; Grigoriadou, S.; Wood, N.W.; Hambleton, S.; Burns, S.O.; Thrasher, A.J.; et al. A Robust Model for Read Count Data in Exome Sequencing Experiments and Implications for Copy Number Variant Calling. *Bioinformatics* **2012**, *28*, 2747–2754. [CrossRef] [PubMed]

86. Krumm, N.; Sudmant, P.H.; Ko, A.; O'Roak, B.J.; Malig, M.; Coe, B.P.; Project, N.E.S.; Quinlan, A.R.; Nickerson, D.A.; Eichler, E.E. Copy Number Variation Detection and Genotyping from Exome Sequence Data. *Genome Res.* **2012**, *22*, 1525–1532. [CrossRef] [PubMed]

87. Shi, Y.; Majewski, J. FishingCNV: A Graphical Software Package for Detecting Rare Copy Number Variations in Exome-Sequencing Data. *Bioinformatics* **2013**, *29*, 1461–1462. [CrossRef] [PubMed]

88. Zhang, Y.; Yu, Z.; Ban, R.; Zhang, H.; Iqbal, F.; Zhao, A.; Li, A.; Shi, Q. DeAnnCNV: A Tool for Online Detection and Annotation of Copy Number Variations from Whole-Exome Sequencing Data. *Nucleic Acids Res.* **2015**, *43*, W289–W294. [CrossRef]

89. Derouault, P.; Chauzeix, J.; Rizzo, D.; Miressi, F.; Magdelaine, C.; Bourthoumieu, S.; Durand, K.; Dzugan, H.; Feuillard, J.; Sturtz, F.; et al. CovCopCan: An Efficient Tool to Detect Copy Number Variation from Amplicon Sequencing Data in Inherited Diseases and Cancer. *PLoS Comput. Biol.* **2020**, *16*, e1007503. [CrossRef]

90. Chanwigoon, S.; Piwluang, S.; Wichadakul, D. InCNV: An Integrated Analysis Tool for Copy Number Variation on Whole Exome Sequencing. *Evol. Bioinforma.* **2020**, *16*, 1176934320956577. [CrossRef]

91. Erikson, G.A.; Deshpande, N.; Kesavan, B.G.; Torkamani, A. SG-ADVISER CNV: Copy-Number Variant Annotation and Interpretation. *Genet. Med.* **2015**, *17*, 714–718. [CrossRef]

92. Chandramohan, R.; Kakkar, N.; Roy, A.; Parsons, D.W. ReconCNV: Interactive Visualization of Copy Number Data from High-Throughput Sequencing. *Bioinformatics* **2021**, *37*, 1164–1167. [CrossRef]

93. Requena, F.; Abdallah, H.H.; García, A.; Nitschké, P.; Romana, S.; Malan, V.; Rausell, A. CNVxplorer: A Web Tool to Assist Clinical Interpretation of CNVs in Rare Disease Patients. *Nucleic Acids Res.* **2021**, *49*, W93–W103. [CrossRef]

94. Gel, B.; Serra, E. KaryoploteR: An R/Bioconductor Package to Plot Customizable Genomes Displaying Arbitrary Data. *Bioinformatics* **2017**, *33*, 3088–3090. [CrossRef] [PubMed]

95. Hahne, F.; Ivanek, R. *Visualizing Genomic Data Using Gviz and Bioconductor BT—Statistical Genomics: Methods and Protocols*; Mathé, E., Davis, S., Eds.; Springer: New York, NY, USA, 2016; pp. 335–351. ISBN 978-1-4939-3578-9.

96. Gel, B.; Magallon, M. CopyNumberPlots: Create Copy-Number Plots Using KaryoploteR Functionality. 2021. Available online: http://github.com/bernatgel/CopyNumberPlots (accessed on 25 October 2022).

97. Shang, X.; Peng, Z.; Ye, Y.; Asan; Zhang, X.; Chen, Y.; Zhu, B.; Cai, W.; Chen, S.; Cai, R.; et al. Rapid Targeted Next-Generation Sequencing Platform for Molecular Screening and Clinical Genotyping in Subjects with Hemoglobinopathies. *EBioMedicine* **2017**, *23*, 150–159. [CrossRef] [PubMed]

98. Chen, P.; Yu, X.; Huang, H.; Zeng, W.; He, X.; Liu, M.; Huang, B. Evaluation of Ion Torrent Next-Generation Sequencing for Thalassemia Diagnosis. *J. Int. Med. Res.* **2020**, *48*, 0300060520967778. [CrossRef] [PubMed]

99. Zhao, J.; Li, J.; Lai, Q.; Yu, Y. Combined Use of Gap-PCR and next-Generation Sequencing Improves Thalassaemia Carrier Screening among Premarital Adults in China. *J. Clin. Pathol.* **2020**, *73*, 488–492. [CrossRef]

100. He, J.; Song, W.; Yang, J.; Lu, S.; Yuan, Y.; Guo, J.; Zhang, J.; Ye, K.; Yang, F.; Long, F.; et al. Next-Generation Sequencing Improves Thalassemia Carrier Screening among Premarital Adults in a High Prevalence Population: The Dai Nationality, China. *Genet. Med.* **2017**, *19*, 1022–1031. [CrossRef]

101. Fan, D.M.; Yang, X.; Huang, L.M.; Ouyang, G.J.; Yang, X.X.; Li, M. Simultaneous Detection of Target CNVs and SNVs of Thalassemia by Multiplex PCR and Next-generation Sequencing. *Mol. Med. Rep.* **2019**, *19*, 2837–2848. [CrossRef]
102. Zebisch, A.; Schulz, E.; Grosso, M.; Lombardo, B.; Acierno, G.; Sill, H.; Iolascon, A. Identification of a Novel Variant of Epsilon-Gamma-Delta-Beta Thalassemia Highlights Limitations of next Generation Sequencing. *Am. J. Hematol.* **2015**, *90*, E52–E54. [CrossRef]
103. Fromer, M.; Moran, J.L.; Chambert, K.; Banks, E.; Bergen, S.E.; Ruderfer, D.M.; Handsaker, R.E.; McCarroll, S.A.; O'Donovan, M.C.; Owen, M.J.; et al. Discovery and Statistical Genotyping of Copy-Number Variation from Whole-Exome Sequencing Depth. *Am. J. Hum. Genet.* **2012**, *91*, 597–607. [CrossRef]
104. Cao, Y.; Chan, G.C.-F. NGS4THAL, a One-Stop Molecular Diagnosis and Carrier Screening Tool for Thalassemia and Other Hemoglobinopathies by next-Generation Sequencing. *Res. Sq.* **2022**, *24*, 1089–1099. [CrossRef]
105. Jiang, F.; Lyu, G.Z.; Zhang, V.W.; Li, D.Z. Identification of Thalassemia Gene Cluster Deletion by Long-Read Whole-Genome Sequencing (LR-WGS). *Int. J. Lab. Hematol.* **2021**, *43*, 859–865. [CrossRef]
106. Li, Q.; Mao, Y.; Li, S.; Du, H.; He, W.; He, J.; Kong, L.; Zhang, J.; Liang, B.; Liu, J. Haplotyping by Linked-Read Sequencing (HLRS) of the Genetic Disease Carriers for Preimplantation Genetic Testing without a Proband or Relatives. *BMC Med. Genomics* **2020**, *13*, 117. [CrossRef] [PubMed]
107. Korlach, J.; Turner, S.W. *Single-Molecule Sequencing BT—Encyclopedia of Biophysics*; Roberts, G.C.K., Ed.; Springer: Berlin/Heidelberg, Germany, 2013; pp. 2344–2347. ISBN 978-3-642-16712-6.
108. Eid, J.; Fehr, A.; Gray, J.; Luong, K.; Lyle, J.; Otto, G.; Peluso, P.; Rank, D.; Baybayan, P.; Bettman, B.; et al. Real-Time DNA Sequencing from Single Polymerase Molecules. *Science* **2009**, *323*, 133–138. [CrossRef] [PubMed]
109. PacBio SMRT Sequencing—How It Works. Available online: https://www.pacb.com/smrt-science/attachment/infographic_smrt-sequencing-how-it-works/%0Awww.pacb.com (accessed on 25 October 2022).
110. Wenger, A.M.; Peluso, P.; Rowell, W.J.; Chang, P.-C.; Hall, R.J.; Concepcion, G.T.; Ebler, J.; Fungtammasan, A.; Kolesnikov, A.; Olson, N.D.; et al. Accurate Circular Consensus Long-Read Sequencing Improves Variant Detection and Assembly of a Human Genome. *Nat. Biotechnol.* **2019**, *37*, 1155–1162. [CrossRef] [PubMed]
111. Wang, Y.; Zhao, Y.; Bollas, A.; Wang, Y.; Au, K.F. Nanopore Sequencing Technology, Bioinformatics and Applications. *Nat. Biotechnol.* **2021**, *39*, 1348–1365. [CrossRef]
112. Tomasz Dobrzycki Selecting the Right Library Prep Method for Your Experiment. Available online: https://nanoporetech.com/resource-centre/video/lc22/selecting-the-right-library-prep-method-for-your-experiment (accessed on 31 October 2022).
113. Oxford Nanopore Technologies How It Works. Available online: https://nanoporetech.com/how-it-works (accessed on 12 October 2022).
114. Carneiro, M.O.; Russ, C.; Ross, M.G.; Gabriel, S.B.; Nusbaum, C.; DePristo, M.A. Pacific Biosciences Sequencing Technology for Genotyping and Variation Discovery in Human Data. *BMC Genomics* **2012**, *13*, 375. [CrossRef]
115. Cheng, S.H.; Jiang, P.; Sun, K.; Cheng, Y.K.Y.; Chan, K.C.A.; Leung, T.Y.; Chiu, R.W.K.; Lo, Y.M.D. Noninvasive Prenatal Testing by Nanopore Sequencing of Maternal Plasma DNA: Feasibility Assessment. *Clin. Chem.* **2015**, *61*, 1305–1306. [CrossRef]
116. Zhang, H.; Jain, C.; Aluru, S. A Comprehensive Evaluation of Long Read Error Correction Methods. *BMC Genomics* **2020**, *21*, 889. [CrossRef]
117. Kolmogorov, M.; Yuan, J.; Lin, Y.; Pevzner, P.A. Assembly of Long, Error-Prone Reads Using Repeat Graphs. *Nat. Biotechnol.* **2019**, *37*, 540–546. [CrossRef]
118. Ruan, J.; Li, H. Fast and Accurate Long-Read Assembly with Wtdbg2. *Nat. Methods* **2020**, *17*, 155–158. [CrossRef]
119. Shafin, K.; Pesout, T.; Lorig-Roach, R.; Haukness, M.; Olsen, H.E.; Bosworth, C.; Armstrong, J.; Tigyi, K.; Maurer, N.; Koren, S.; et al. Nanopore Sequencing and the Shasta Toolkit Enable Efficient de Novo Assembly of Eleven Human Genomes. *Nat. Biotechnol.* **2020**, *38*, 1044–1053. [CrossRef]
120. Morisse, P.; Marchet, C.; Limasset, A.; Lecroq, T.; Lefebvre, A. Scalable Long Read Self-Correction and Assembly Polishing with Multiple Sequence Alignment. *Sci. Rep.* **2021**, *11*, 761. [CrossRef] [PubMed]
121. Xiao, C.-L.; Chen, Y.; Xie, S.-Q.; Chen, K.-N.; Wang, Y.; Han, Y.; Luo, F.; Xie, Z. MECAT: Fast Mapping, Error Correction, and de Novo Assembly for Single-Molecule Sequencing Reads. *Nat. Methods* **2017**, *14*, 1072–1074. [CrossRef] [PubMed]
122. Koren, S.; Walenz, B.P.; Berlin, K.; Miller, J.R.; Bergman, N.H.; Phillippy, A.M. Canu: Scalable and Accurate Long-Read Assembly via Adaptive k-Mer Weighting and Repeat Separation. *Genome Res.* **2017**, *27*, 722–736. [CrossRef] [PubMed]
123. Berlin, K.; Koren, S.; Chin, C.-S.; Drake, J.P.; Landolin, J.M.; Phillippy, A.M. Assembling Large Genomes with Single-Molecule Sequencing and Locality-Sensitive Hashing. *Nat. Biotechnol.* **2015**, *33*, 623–630. [CrossRef] [PubMed]
124. Chin, C.-S.; Peluso, P.; Sedlazeck, F.J.; Nattestad, M.; Concepcion, G.T.; Clum, A.; Dunn, C.; O'Malley, R.; Figueroa-Balderas, R.; Morales-Cruz, A.; et al. Phased Diploid Genome Assembly with Single-Molecule Real-Time Sequencing. *Nat. Methods* **2016**, *13*, 1050–1054. [CrossRef]
125. Chaisson, M.J.; Tesler, G. Mapping Single Molecule Sequencing Reads Using Basic Local Alignment with Successive Refinement (BLASR): Application and Theory. *BMC Bioinformatics* **2012**, *13*, 238. [CrossRef]
126. Chen, Y.; Nie, F.; Xie, S.-Q.; Zheng, Y.-F.; Dai, Q.; Bray, T.; Wang, Y.-X.; Xing, J.-F.; Huang, Z.-J.; Wang, D.-P.; et al. Efficient Assembly of Nanopore Reads via Highly Accurate and Intact Error Correction. *Nat. Commun.* **2021**, *12*, 60. [CrossRef]

127. Shafin, K.; Pesout, T.; Chang, P.-C.; Nattestad, M.; Kolesnikov, A.; Goel, S.; Baid, G.; Kolmogorov, M.; Eizenga, J.M.; Miga, K.H.; et al. Haplotype-Aware Variant Calling with PEPPER-Margin-DeepVariant Enables High Accuracy in Nanopore Long-Reads. *Nat. Methods* **2021**, *18*, 1322–1332. [CrossRef]
128. Ahsan, M.U.; Liu, Q.; Fang, L.; Wang, K. NanoCaller for Accurate Detection of SNPs and Indels in Difficult-to-Map Regions from Long-Read Sequencing by Haplotype-Aware Deep Neural Networks. *Genome Biol.* **2021**, *22*, 261. [CrossRef]
129. Su, J.; Zheng, Z.; Ahmed, S.S.; Lam, T.-W.; Luo, R. Clair3-Trio: High-Performance Nanopore Long-Read Variant Calling in Family Trios with Trio-to-Trio Deep Neural Networks. *Brief. Bioinform.* **2022**, bbac301. [CrossRef]
130. Ramachandran, A.; Lumetta, S.S.; Klee, E.W.; Chen, D. HELLO: Improved Neural Network Architectures and Methodologies for Small Variant Calling. *BMC Bioinformatics* **2021**, *22*, 404. [CrossRef] [PubMed]
131. Smolka, M.; Paulin, L.F.; Grochowski, C.M.; Mahmoud, M.; Behera, S.; Gandhi, M.; Hong, K.; Pehlivan, D.; Scholz, S.W.; Carvalho, C.M.B.; et al. Comprehensive Structural Variant Detection: From Mosaic to Population-Level (Sniffles2). *bioRxiv* **2022**. [CrossRef]
132. Cao, S.; Jiang, T.; Liu, Y.; Liu, S.; Wang, Y. Re-Genotyping Structural Variants through an Accurate Force-Calling Method. *bioRxiv* **2022**. [CrossRef]
133. Heller, D.; Vingron, M. SVIM: Structural Variant Identification Using Mapped Long Reads. *Bioinformatics* **2019**, *35*, 2907–2915. [CrossRef]
134. Tham, C.Y.; Tirado-Magallanes, R.; Goh, Y.; Fullwood, M.J.; Koh, B.T.H.; Wang, W.; Ng, C.H.; Chng, W.J.; Thiery, A.; Tenen, D.G.; et al. NanoVar: Accurate Characterization of Patients' Genomic Structural Variants Using Low-Depth Nanopore Sequencing. *Genome Biol.* **2020**, *21*, 56. [CrossRef]
135. Xu, L.; Mao, A.; Liu, H.; Gui, B.; Choy, K.W.; Huang, H.; Yu, Q.; Zhang, X.; Chen, M.; Lin, N.; et al. Long-Molecule Sequencing: A New Approach for Identification of Clinically Significant DNA Variants in α-Thalassemia and β-Thalassemia Carriers. *J. Mol. Diagnostics* **2020**, *22*, 1087–1095. [CrossRef] [PubMed]
136. Liang, Q.; Gu, W.; Chen, P.; Li, Y.; Liu, Y.; Tian, M.; Zhou, Q.; Qi, H.; Zhang, Y.; He, J.; et al. A More Universal Approach to Comprehensive Analysis of Thalassemia Alleles (CATSA). *J. Mol. Diagnostics* **2021**, *23*, 1195–1204. [CrossRef]
137. Peng, C.; Zhang, H.; Ren, J.; Chen, H.; Du, Z.; Zhao, T.; Mao, A.; Xu, R.; Lu, Y.; Wang, H.; et al. Analysis of Rare Thalassemia Genetic Variants Based on Third-Generation Sequencing. *Sci. Rep.* **2022**, *12*, 9907. [CrossRef]
138. Li, Y.; Liang, L.; Qin, T.; Tian, M. Detection of Hemoglobin H Disease by Long Molecule Sequencing. *J. Clin. Lab. Anal.* **2022**, *36*, e24687. [CrossRef]
139. Jiang, F.; Liu, W.; Zhang, L.; Guo, Y.; Chen, M.; Zeng, X.; Wang, Y.; Li, Y.; Xian, J.; Du, B.; et al. Noninvasive Prenatal Testing for β-Thalassemia by Targeted Nanopore Sequencing Combined with Relative Haplotype Dosage (RHDO): A Feasibility Study. *Sci. Rep.* **2021**, *11*, 5714. [CrossRef]
140. Christopher, H.; Burns, A.; Josephat, E.; Makani, J.; Schuh, A.; Nkya, S. Using DNA Testing for the Precise, Definite, and Low-Cost Diagnosis of Sickle Cell Disease and Other Haemoglobinopathies: Findings from Tanzania. *BMC Genomics* **2021**, *22*, 902. [CrossRef]
141. Liu, S.; Wang, H.; Leigh, D.; Cram, D.S.; Wang, L.; Yao, Y. Third-Generation Sequencing: Any Future Opportunities for PGT? *J. Assist. Reprod. Genet.* **2021**, *38*, 357–364. [CrossRef]
142. Sirén, J.; Monlong, J.; Chang, X.; Novak, A.M.; Eizenga, J.M.; Markello, C.; Sibbesen, J.A.; Hickey, G.; Chang, P.-C.; Carroll, A.; et al. Pangenomics Enables Genotyping of Known Structural Variants in 5202 Diverse Genomes. *Science* **2021**, *374*, abg8871. [CrossRef] [PubMed]
143. Li, H.; Feng, X.; Chu, C. The Design and Construction of Reference Pangenome Graphs with Minigraph. *Genome Biol.* **2020**, *21*, 265. [CrossRef] [PubMed]
144. Wang, T.; Antonacci-Fulton, L.; Howe, K.; Lawson, H.A.; Lucas, J.K.; Phillippy, A.M.; Popejoy, A.B.; Asri, M.; Carson, C.; Chaisson, M.J.P.; et al. The Human Pangenome Project: A Global Resource to Map Genomic Diversity. *Nature* **2022**, *604*, 437–446. [CrossRef] [PubMed]
145. Roy, S.; Coldren, C.; Karunamurthy, A.; Kip, N.S.; Klee, E.W.; Lincoln, S.E.; Leon, A.; Pullambhatla, M.; Temple-Smolkin, R.L.; Voelkerding, K.V.; et al. Standards and Guidelines for Validating Next-Generation Sequencing Bioinformatics Pipelines: A Joint Recommendation of the Association for Molecular Pathology and the College of American Pathologists. *J. Mol. Diagnostics* **2018**, *20*, 4–27. [CrossRef]
146. Santani, A.; Simen, B.B.; Briggs, M.; Lebo, M.; Merker, J.D.; Nikiforova, M.; Vasalos, P.; Voelkerding, K.; Pfeifer, J.; Funke, B. Designing and Implementing NGS Tests for Inherited Disorders: A Practical Framework with Step-by-Step Guidance for Clinical Laboratories. *J. Mol. Diagnostics* **2019**, *21*, 369–374. [CrossRef]
147. Roy, N.B.A.; Da Costa, L.; Russo, R.; Bianchi, P.; del Mar Mañú-Pereira, M.; Fermo, E.; Andolfo, I.; Clark, B.; Proven, M.; Sanchez, M.; et al. The Use of Next-Generation Sequencing in the Diagnosis of Rare Inherited Anaemias: A Joint BSH/EHA Good Practice Paper. *HemaSphere* **2022**, *6*, e739. [CrossRef]
148. International Organization for Standardization [ISO] ISO 20397-1:2022 Biotechnology—Massively Parallel Sequencing. Available online: https://www.iso.org/standard/74054.html (accessed on 14 November 2022).