Genomic Characterization of Rare Pathogenic Variants in Sickle Cell Disease Patients Using Oxford Nano pore Sequencing: Current Advances, Clinical Applications, and Future Perspectives

Isyaku Alhassan[1], Nura Garba[2], Muhammad Yalwa Gwarzo[3], Muhammad Bashir Tijjani[4]

[1] Department of Medical Laboratory Science, Maryam Abacha American University of Niger, Nigeria
[2] Department of Medical Laboratory Science (Hematology Unit), Bayero University Kano, Nigeria
[3] Department of Medical Laboratory Science (Chemical Pathology Unit), Bayero University, Nigeria
[4] Department of Microbiology, Ahmadu Bello University, Nigeria

ABSTRACT

Sickle cell disease (SCD) is one of the most prevalent monogenic hereditary disorders worldwide and remains a major public health challenge, particularly in sub-Saharan Africa, the Middle East, India, and the Mediterranean region. Although SCD results from a single nucleotide substitution (c.20A>T) in the HBB gene causing the Glu6Val alteration in β-globin, considerable variability exists in disease severity, complications, and therapeutic responses among individuals with identical genotypes. This phenotypic diversity is influenced by additional genetic and epigenetic factors, including rare pathogenic variants, structural genomic alterations, copy number variations, and modifier genes involved in fatal haemoglobin regulation, inflammation, oxidative stress, erythropoiesis, and vascular biology. Conventional sequencing approaches have improved molecular diagnosis but remain limited in resolving complex genomic regions, structural variants, repetitive sequences, and long-range haplotypes. Oxford Nano pore Sequencing (ONS), a third-generation long-read sequencing technology, provides a transformative approach by enabling real-time sequencing of native DNA molecules, generation of ultra-long reads, accurate haplotype phasing, detection of structural variants, and simultaneous identification of DNA methylation patterns. This review examines the role of ONS in characterizing rare pathogenic variants associated with SCD heterogeneity and evaluates its applications in precision medicine. It discusses advances in long-read genomic analysis, detection of disease-modifying variants, characterization of the β-globin locus, pharmacogenomics profiling, and monitoring of emerging gene-editing therapies. The review also highlights current challenges, including sequencing accuracy, bioinformatics standardization, ethical considerations, cost, and limited representation of diverse populations in genomic databases. As Nano pore technology continues to improve, it holds significant potential to enhance diagnosis, prognosis, therapeutic selection, and personalized management of sickle cell disease.

KEYWORDS

Sickle Cell Disease, Oxford Nano pore Sequencing, Long-read Sequencing, Rare Pathogenic Variants, Precision Medicine, Structural Variants, Genomic Medicine, HBB Gene, Genetic Modifiers, Fetal Haemoglobin, Copy Number Variations, Haplotype Phasing, Epigenetics, Clinical Genomics, Personalized Medicine

I. INTRODUCTION

Sickle cell disease (SCD) is one of the most extensively investigated monogenic disorders and remains a major global health challenge despite decades of advances in molecular biology, clinical management, and therapeutic interventions. The disease affects millions of individuals worldwide and is particularly prevalent in sub-Saharan Africa, the Middle East, India, and parts of the Mediterranean, where the sickle cell allele has historically been maintained because of its protective effect against severe Plasmodium falciparum malaria. Each year, an estimated 300,000–400,000 infants are born with SCD globally, with projections indicating a continued increase due to population growth and improved childhood survival. Although significant progress has been made in new-born screening, supportive care, hydro xyurea therapy, hematopoietic stem cell transplantation, and emerging gene-editing approaches, SCD continues to impose substantial morbidity, mortality, and socioeconomic burdens, particularly in low- and middle-income countries where access to specialized healthcare and genomic diagnostics remains limited.

The molecular basis of sickle cell disease is a single nucleotide substitution (c.20A>T) in the β-globin (HBB) gene located on the short arm of chromosome 11 (11p15.5). This point mutation replaces adenine with thymine in the sixth codon of the HBB gene, resulting in the substitution of glutamic acid by valise at the sixth amino acid position of the β-globin polypeptide chain (Glu6Val). Consequently, the abnormal haemoglobin variant known as haemoglobin S (HbS) is produced instead of normal adult haemoglobin (HbA). Although this genetic alteration involves only a single nucleotide, it profoundly changes the physicochemical properties of haemoglobin, initiating a cascade of molecular and cellular events responsible for the pathophysiology of the disease.

Under conditions of reduced oxygen tension, dehydration, acidosis, or oxidative stress, HbS molecules undergo reversible polymerization, forming long intracellular fibres that distort erythrocytes into the characteristic sickle or crescent shape. Repeated cycles of deoxygenating and oxygenation progressively damage the red blood cell membrane, resulting in reduced cellular deformability, increased membrane rigidity, altered ion transport, and enhanced adhesion to vascular endothelial cells. These pathological changes shorten erythrocyte lifespan from approximately 120 days to 10–20 days, leading to chronic intravascular and extravascular haemolysis. Simultaneously, rigid sickled erythrocytes obstruct the microcirculation, producing recurrent vaso-occlusive episodes that impair tissue perfusion and cause ischemic injury in multiple organs. The resulting inflammatory response, endothelial dysfunction, nitric oxide depletion, oxidative stress, platelet activation, and coagulation abnormalities collectively contribute to the complex multisystem manifestations of SCD.

Clinically, sickle cell disease is characterized by remarkable heterogeneity despite its apparently simple monogenic origin. While all individuals with homozygous sickle cell anaemia (HbSS) harbour the same pathogenic HBB mutation, disease severity varies considerably among affected patients. Some individuals experience relatively mild anaemia and infrequent pain crises, allowing near-normal quality of life, whereas others develop recurrent vaso-occlusive crises, acute chest syndrome, stroke, pulmonary hypertension, chronic kidney disease, retinopathy, avascular necrosis, leg ulcers, splenic dysfunction, severe infections, and premature death despite receiving similar clinical care. This wide

spectrum of clinical outcomes demonstrates that the primary HBB mutation alone is insufficient to explain the observed phenotypic diversity, indicating that additional genetic, epigenetic, and environmental factors play critical roles in determining disease expression.

Over the past two decades, advances in human genomics have revealed that numerous genetic modifiers contribute significantly to the variability observed in SCD. Variants influencing fetal haemoglobin (HbF) production represent some of the most important disease modifiers because elevated HbF inhibits HbS polymerization and reduces erythrocyte sickling. Genome-wide association studies have identified major quantitative trait loci within BCL11A, HBS1L-MYB, and the β-globin gene cluster that regulate HbF expression and are associated with reduced disease severity. In addition, rare pathogenic variants affecting genes involved in erythropoiesis, inflammation, oxidative stress, endothelial function, iron metabolism, immune regulation, and renal physiology have been linked to differences in clinical phenotype, susceptibility to complications, and treatment response. These discoveries have shifted the understanding of SCD from a simple monogenic disorder toward a genetically complex disease influenced by multiple interacting molecular pathways.

Epigenetic mechanisms further contribute to disease heterogeneity by regulating gene expression without altering the underlying DNA sequence. DNA methylation, histone modifications, chromatin remodelling, and non-coding RNAs influence elytroid differentiation, globin gene switching, inflammatory signalling, and vascular homeostasis. Variations in these regulatory processes may partly explain why patients with identical genotypes exhibit markedly different clinical manifestations and therapeutic responses. Increasing evidence suggests that integrating genomic and epigenetic information will improve understanding of disease mechanisms and facilitate the development of individualized therapeutic strategies for SCD.

The recognition that genetic and epigenetic modifiers substantially influence disease progression has increased the demand for comprehensive genomic technologies capable of accurately identifying rare pathogenic variants, structural rearrangements, copy number variations, repeat expansions, and long-range haplotypes. Conventional molecular diagnostic techniques, including Sanger sequencing and short-read next-generation sequencing (NGS), have significantly advanced the molecular diagnosis of SCD but remain limited in resolving repetitive genomic regions, homologous sequences within the β-globin locus, and complex structural variants. These limitations have stimulated the adoption of long-read sequencing technologies, particularly Oxford Nano pore Sequencing (ONS), which enables real-time sequencing of native DNA molecules while simultaneously detecting epigenetic modifications and resolving complex genomic architectures. The application of ONS is therefore emerging as a powerful tool for elucidating the molecular basis of phenotypic variability, improving molecular diagnosis, identifying novel disease-modifying variants, and advancing precision medicine for patients with sickle cell disease.

This review aims to provide a comprehensive overview of the genomic architecture of sickle cell disease with particular emphasis on the role of rare pathogenic variants in determining clinical heterogeneity. It further examines recent advances in Oxford Nano pore sequencing technology, evaluates its applications in the identification of disease-modifying variants and structural genomic alterations, discusses current analytical workflows and bioinformatics approaches, and explores its potential integration into precision medicine. Finally, the review

highlights existing challenges, ethical considerations, and future perspectives for incorporating long-read sequencing into routine clinical care to improve diagnosis, prognostication, and personalized management of individuals living with sickle cell disease.

For several decades, investigations into the genetic basis of clinical variability in sickle cell disease (SCD) focused predominantly on common genetic modifiers that regulate fetal haemoglobin (HbF) production. Among the most extensively studied are variants within BCL11A, the HBS1L-MYB intragenic region, and KLF1, all of which play central roles in elytroid differentiation and the developmental switching of globin gene expression. Elevated HbF concentrations are strongly associated with reduced disease severity because HbF interferes with the polymerization of deoxygenated haemoglobin S (HbS), thereby decreasing erythrocyte sickling, haemolysis, and vaso-occlusive events. Genome-wide association studies have consistently demonstrated that polymorphisms within these loci account for a substantial proportion of the inter individual variation in HbF levels and are associated with improved clinical outcomes, including reduced frequency of pain crises, lower risk of stroke, and enhanced overall survival. These discoveries have significantly advanced the understanding of disease modulation and have contributed to the development of therapeutic strategies aimed at pharmacologically inducing HbF expression, such as hydro urea therapy.

Despite these important advances, the known HbF-associated modifier loci explain only a portion of the remarkable phenotypic diversity observed among patients with SCD. Increasing evidence from whole-genome sequencing, whole-exome sequencing, and long-read genomic studies indicates that numerous rare pathogenic variants independently influence disease severity through diverse biological pathways beyond globin gene regulation. Variants affecting genes involved in inflammatory signalling, oxidative stress responses, endothelial activation, nitric oxide metabolism, coagulation cascades, complement activation, iron homeostasis, immune regulation, apoptosis, and DNA damage repair have been associated with susceptibility to severe clinical complications, including acute chest syndrome, stroke, pulmonary hypertension, nephropathy, and chronic organ dysfunction. Because many of these variants occur at low frequencies within populations, they have historically been underrepresented or completely overlooked in conventional genetic association studies. Their cumulative contribution, however, is increasingly recognized as an important determinant of individualized disease expression, prognosis, and therapeutic responsiveness.

In addition to single nucleotide variants, structural genomic alterations—including copy number variations (CNVs), insertions, deletions, inversions, translocations, repeat expansions, and complex chromosomal rearrangements—are emerging as critical contributors to SCD pathogenesis and clinical heterogeneity. These genomic alterations can disrupt coding sequences, modify regulatory elements, alter gene dosage, or influence long-range chromatin interactions, ultimately affecting gene expression and cellular function. Furthermore, the interaction between rare pathogenic variants and environmental factors, such as recurrent infections, nutritional status, hypoxia, and healthcare access, creates highly individualized disease trajectories that cannot be fully explained by the presence of the pathogenic HBB mutation alone. This increasingly sophisticated understanding has shifted the paradigm of SCD from a classic single-gene disorder to a genetically and biologically complex disease influenced by multiple interacting molecular mechanisms.

The rapid evolution of genomic sequencing technologies has dramatically enhanced the ability to detect these previously hidden contributors to disease variability. Although Sanger sequencing and second-generation short-read next-generation sequencing (NGS) have transformed molecular diagnostics by enabling accurate identification of single nucleotide variants and small insertions or deletions, these technologies remain limited in their capacity to resolve repetitive genomic regions, highly homologous sequences, structural variants, and long-range haplotypes. Fragmentation of genomic DNA into relatively short sequencing reads often results in incomplete reconstruction of complex genomic loci, reducing the sensitivity for detecting clinically significant structural alterations and limiting comprehensive genomic interpretation. These limitations have created a growing demand for sequencing technologies capable of providing continuous, high-resolution characterization of entire genomic regions.

Among recent technological innovations, Oxford Nano pore Sequencing (ONS) has emerged as one of the most transformative platforms in genomic medicine. Unlike sequencing-by-synthesis technologies, ONS directly sequences native DNA and RNA molecules as they pass through biological Nano pores, generating exceptionally long sequencing reads without the need for polymerase chain reaction (PCR) amplification. Preservation of long DNA fragments enables accurate reconstruction of complex genomic architectures, direct phasing of maternal and paternal haplotypes, comprehensive identification of structural variants, and simultaneous detection of epigenetic modifications such as DNA methylation. The ability to analyse genomic sequence and epigenetic information within a single experiment represents a significant advancement over conventional sequencing methods and provides unprecedented opportunities for investigating the molecular mechanisms underlying inherited diseases.

These technological advantages make Oxford Nano pore Sequencing particularly valuable for studying the β-globin gene cluster and its surrounding regulatory regions, where structural complexity, repetitive sequences, and homologous genomic elements often complicate analysis using short-read sequencing platforms. Long-read sequencing facilitates precise characterization of the primary pathogenic HBB mutation together with neighbouring modifier alleles, structural rearrangements, and copy number changes, recombination events, and phased haplotypes that collectively influence disease phenotype. Moreover, the technology enables identification of rare pathogenic variants that may modify fetal haemoglobin regulation, inflammatory pathways, vascular biology, oxidative stress responses, erythrocyte membrane stability, and immune function. Such comprehensive genomic characterization provides deeper insights into the biological mechanisms responsible for interp atient variability and supports the development of more accurate molecular classification systems for SCD.

Consequently, Oxford Nano pore Sequencing represents a highly promising platform for advancing precision medicine in sickle cell disease. By integrating analyses of the primary disease-causing mutation with rare pathogenic variants, structural genomic alterations, haplotype architecture, copy number variation, and epigenetic modifications, Nano pore sequencing offers a comprehensive understanding of each patient's unique genomic profile. This integrated approach has important implications for improving molecular diagnosis, predicting disease severity, identifying individuals at increased risk of specific complications, optimizing pharmacogenomics-guided treatment strategies, monitoring responses to

therapies such as hydroxyl urea, and selecting suitable candidates for emerging gene-editing and gene-replacement interventions. As sequencing chemistry, bioinformatics algorithms, and clinical interpretation frameworks continue to improve, Oxford Nanopore Sequencing is expected to become an increasingly important component of precision haematology, facilitating individualized management strategies and improving long-term outcomes for patients living with sickle cell disease.

II. MOLECULAR GENETICS OF SICKLE CELL DISEASE

Sickle cell disease (SCD) is one of the best-characterized monogenic disorders in human genetics, yet its molecular pathogenesis is considerably more complex than would be expected from a single-gene mutation. The disease originates from a single nucleotide substitution (c.20A>T) in exon 1 of the HBB gene located on chromosome 11p15.5. This point mutation changes the sixth codon from GAG to GTG, replacing the hydrophilic amino acid glutamic acid with the hydrophobic amino acid valine at position six of the β-globin protein (Glu6Val). The resulting abnormal β-globin chain combines with α-globin chains to form haemoglobin S (HbS), the hallmark molecular defect responsible for SCD. Although this mutation represents only one nucleotide change within the approximately 3.2-billion-base-pair human genome, it fundamentally alters haemoglobin structure and initiates a cascade of pathological events that affect nearly every organ system.

The substitution of glutamic acid by valine creates a hydrophobic region on the surface of the HbS molecule. Under physiological oxygenated conditions, HbS behaves similarly to normal adult haemoglobin (HbA). However, during deoxygenating, dehydration, acidosis, or oxidative stress, the exposed hydrophobic valine residue promotes abnormal intermolecular interactions between adjacent HbS molecules. These interactions trigger the polymerization of deoxygenated HbS into long, rigid intracellular fibres that distort erythrocytes into the characteristic crescent or sickle shape. Initially, sickling is reversible upon oxygenation, but repeated cycles of polymerization and depolymerisation progressively damage the erythrocyte membrane, ultimately producing irreversibly sickled cells with severely compromised structural integrity and function.

Polymerization of HbS profoundly alters the mechanical and biochemical properties of red blood cells. Sickled erythrocytes exhibit markedly reduced deformability, increased membrane rigidity, abnormal ion transport, dehydration, oxidative membrane injury, phosphatide serine externalization, and shortened cellular survival. Instead of surviving approximately 120 days like normal erythrocytes, sickled red blood cells typically persist for only 10–20 days before undergoing premature destruction. Their reduced flexibility impairs passage through the microvasculature, while increased expression of adhesion molecules promotes interactions with endothelial cells, leukocytes, platelets, and the extracellular matrix. These abnormalities collectively contribute to recurrent micro vascular obstruction and tissue ischemia, which are central pathological features of SCD.

Vaso-occlusion is now recognized as a complex multicellular process rather than a simple mechanical blockage caused by rigid erythrocytes. Sickled red blood cells adhere to activated vascular endothelial cells through multiple adhesion molecules, including selections, integrin’s, and immunoglobulin superfamily proteins. Simultaneously, activated

neutrophils, monocytes, platelets, and circulating extracellular vesicles interact with erythrocytes to form multicellular aggregates that further obstruct blood flow. Endothelial activation stimulates the expression of inflammatory cytokines, chemokine's, and adhesion receptors, amplifying leukocyte recruitment and sustaining chronic vascular inflammation. This inflammatory microenvironment perpetuates recurrent vaso-occlusive crises, resulting in severe pain episodes and progressive tissue injury throughout the body.

Chronic haemolysis constitutes another fundamental mechanism contributing to disease pathophysiology. Continuous destruction of sickled erythrocytes releases free haemoglobin, home, and erythrocyte-derived micro particles into the circulation. Cell-free haemoglobin rapidly scavenges nitric oxide (NO), a critical vasodilator that maintains vascular homeostasis by regulating endothelial function, platelet activation, and smooth muscle relaxation. Reduced NO bioavailability promotes vasoconstriction, endothelial dysfunction, platelet activation, oxidative stress, and vascular remodelling. Free home further activates innate immune pathways through Toll-like receptor 4 (TLR4), stimulates complement activation, and enhances production of reactive oxygen species, thereby intensifying systemic inflammation and endothelial injury. These processes contribute significantly to the development of pulmonary hypertension, nephropathy, cerebrovascular disease, and other chronic vascular complications associated with SCD.

Repeated episodes of ischemia followed by reperfusion exacerbate tissue injury through the generation of reactive oxygen species, mitochondrial dysfunction, inflammatory mediator release, and activation of programmed cell death pathways. Over time, cumulative ischemia–reperfusion injury leads to progressive dysfunction of multiple organs, including the kidneys, lungs, liver, spleen, retina, heart, bones, and central nervous system. Consequently, SCD is no longer regarded solely as a disorder of abnormal erythrocytes but rather as a chronic systemic disease involving complex interactions among hematologic, vascular, inflammatory, immunologic, and metabolic pathways. This broader understanding has fundamentally reshaped approaches to disease management and therapeutic development.

Despite sharing the same pathogenic HBB mutation, individuals with homozygous HbSS disease display striking differences in clinical severity, age at onset of complications, frequency of vaso-occlusive crises, response to therapy, and long-term survival. This variability underscores the importance of genomic modifiers that influence disease expression beyond the primary mutation. Genome-wide association studies (GWAS) have identified several common modifier loci, including BCL11A, the HBS1L-MYB intragenic region, and polymorphisms within the β-globin gene cluster, which regulate fatal haemoglobin (HbF) production and partially explain differences in clinical phenotype. Nevertheless, these common variants account for only a proportion of the observed inter individual variability, suggesting that additional molecular determinants remain undiscovered.

Increasing attention is therefore being directed toward rare pathogenic variants distributed throughout the genome. These variants occur in genes involved in erythropoiesis, inflammatory signalling, oxidative stress responses, endothelial biology, coagulation, complement activation, iron metabolism, immune regulation, apoptosis, DNA repair, and cellular stress responses. Although individually uncommon, rare coding mutations may exert

substantial biological effects and contribute to severe disease phenotypes, treatment resistance, or susceptibility to specific complications. In parallel, non-coding regulatory variants located within promoters, enhancers, silencers, untranslated regions, and long non-coding RNAs are increasingly recognized as important modulators of gene expression. Structural genomic alterations—including copy number variations, insertions, deletions, inversions, translocations, and repeat expansions—as well as epigenetic modifications such as DNA methylation and chromatin remodelling further contribute to the molecular complexity of SCD. Many of these genomic features remain difficult to detect using conventional short-read sequencing technologies, highlighting the growing importance of long-read sequencing platforms such as Oxford Nano pore sequencing for comprehensive genomic characterization and precision medicine applications

III. RARE PATHOGENIC VARIANTS IN SICKLE CELL DISEASE

Rare pathogenic variants are increasingly recognized as important contributors to the molecular complexity and clinical heterogeneity of sickle cell disease (SCD). In human genetics, rare variants are generally defined as genetic alterations with a minor allele frequency (MAF) of less than 1% within a given population. Although these variants occur infrequently, they often have larger biological effect sizes than common polymorphisms because they can substantially alter protein structure, gene regulation, or cellular signalling pathways. Advances in high-throughput genomic technologies have demonstrated that many rare variants are clinically relevant, influencing disease susceptibility, severity, therapeutic response, and long-term prognosis. In SCD, accumulating evidence indicates that these variants help explain why patients with the same pathogenic HBB mutation frequently experience markedly different clinical outcomes.

Unlike common genetic modifiers identified through genome-wide association studies (GWAS), rare pathogenic variants are often unique to individuals, families, or specific populations and therefore require comprehensive sequencing approaches for accurate detection. Many of these variants have escaped identification because traditional genotyping arrays and short-read sequencing technologies are primarily optimized to detect common single nucleotide polymorphisms (SNPs). The increasing application of whole-genome sequencing, whole-exam sequencing, and long-read sequencing technologies has substantially expanded the discovery of rare coding and non-coding variants that contribute to disease variability in SCD. These findings support the concept that SCD is influenced not only by its primary monogenic defect but also by a diverse network of secondary genomic alterations that collectively modify disease phenotype.

Rare pathogenic variants in SCD encompass multiple classes of genomic alterations, each capable of affecting gene expression or protein function through distinct molecular mechanisms. Missense mutations result in single amino acid substitutions that may alter protein folding, stability, enzymatic activity, receptor binding, or intracellular signalling. Depending on the affected gene and the functional importance of the altered residue, these mutations may enhance susceptibility to inflammation, endothelial dysfunction, oxidative stress, or vascular injury. Nonsense mutations, which introduce premature termination codons, can produce truncated proteins or trigger nonsense-mediated mRNA decay, resulting in partial or complete loss of gene function. Frame shift insertions and deletions

(idles) disrupt the normal reading frame of protein-coding genes, frequently generating severely altered or non-functional proteins with profound biological consequences.

Another important category comprises splice-site variants, which interfere with normal pre-messenger RNA (pre-mRNA) splicing by disrupting conserved donor or acceptor splice sequences. These variants may lead to exon skipping, intron retention, or the generation of abnormal transcript isoforms, ultimately altering protein structure and function. Because precise RNA splicing is essential for normal erythropoiesis, immune regulation, and vascular homeostasis, splice-site mutations can significantly influence disease severity and contribute to inter individual variability in clinical manifestations.

Structural genomic alterations represent an additional class of rare pathogenic variants with increasing importance in SCD research. These include large insertions, deletions, inversions, duplications, translocations, and complex chromosomal rearrangements that modify genomic architecture without necessarily changing individual nucleotide sequences. Structural variants may disrupt coding regions, reposition regulatory elements, alter chromatin organization, or affect long-range interactions between genes and enhancers. Similarly, copy number variations (CNVs), involving gains or losses of genomic segments, can influence gene dosage and consequently alter protein expression levels. Although many structural variants remain undetected by conventional short-read sequencing platforms, long-read sequencing technologies have greatly improved their identification and characterization.

Another emerging category includes repeat expansions, in which repetitive DNA sequences undergo abnormal expansion beyond their normal length. While repeat expansion disorders are classically associated with neurological diseases, increasing evidence suggests that unstable repetitive elements can influence chromatin organization, gene regulation, genomic stability, and transcriptional activity in a variety of inherited disorders. In addition, regulatory mutations occurring within promoters, enhancers, silencers, untranslated regions (UTRs), and long non-coding RNAs may profoundly alter gene expression without changing protein-coding sequences. Such variants are particularly important because they can influence elytroid differentiation, fatal haemoglobin production, inflammatory responses, endothelial activation, and oxidative stress pathways that contribute to disease progression in SCD.

Recent genomic investigations have identified rare pathogenic variants in numerous genes involved in vascular biology, inflammatory signalling, coagulation pathways, complement activation, oxidative stress, renal physiology, immune regulation, cellular adhesion, and nitric oxide metabolism. Variants affecting these biological pathways have been associated with increased susceptibility to severe complications such as ischemic stroke, chronic kidney disease, pulmonary hypertension, recurrent vaso-occlusive crises, acute chest syndrome, avascular necrosis, priapism, leg ulcers, and early mortality. For example, variants in genes regulating endothelial function and vascular integrity may increase the risk of vaso-occlusion, whereas alterations affecting inflammatory mediators or complement proteins may amplify chronic systemic inflammation. Similarly, mutations influencing oxidative stress responses can accelerate erythrocyte damage and haemolysis, thereby worsening disease severity. Although many of these associations require further validation in larger and more

diverse populations, they illustrate the broad genomic landscape that contributes to the complexity of SCD.

One of the most significant observations emerging from recent genomic studies is that patients who share identical homozygous HbSS genotypes frequently possess entirely different combinations of rare modifier variants distributed throughout the genome. Consequently, two individuals carrying the same disease-causing HBB mutation may exhibit dramatically different clinical courses, therapeutic responses, and long-term prognoses because of differences in their broader genomic backgrounds. This concept of genetic individuality provides a compelling explanation for the unpredictable nature of SCD and highlights the limitations of relying solely on the primary pathogenic mutation for disease classification or prognostic assessment. Comprehensive characterization of each patient's complete genomic profile, including both common and rare variants, is therefore becoming increasingly important for individualized risk prediction and precision medicine.

The identification and interpretation of rare pathogenic variants have been greatly accelerated by the development of long-read sequencing technologies, particularly Oxford Nano pore Sequencing (ONS). Unlike conventional short-read platforms, ONS enables the simultaneous detection of single nucleotide variants, insertions, deletions, structural variants, copy number alterations, repeat expansions, phased haplotypes, and epigenetic modifications within a single sequencing experiment. This comprehensive genomic resolution is particularly valuable for uncovering previously undetected modifier variants that influence disease severity and treatment outcomes in SCD. As analytical pipelines, variant interpretation frameworks, and population reference databases continue to improve, integrating rare variant analysis into routine clinical practice has the potential to transform diagnosis, prognostic assessment, and personalized therapeutic decision-making for individuals living with sickle cell disease.

IV. OXFORD NANOPORE SEQUENCING TECHNOLOGY

Oxford Nano pore Sequencing (ONS) represents one of the most significant technological advances in modern genomics and has fundamentally transformed the way genetic information is analysed. As a third-generation sequencing platform, ONS differs substantially from first- and second-generation sequencing technologies by enabling the direct sequencing of native DNA and RNA molecules in real time without requiring sequencing-by-synthesis or amplification-based detection. Rather than relying on fluorescently labelled nucleotides or optical imaging systems, Nano pore sequencing identifies nucleotide sequences by monitoring changes in electrical current as individual nucleic acid molecules pass through biological Nano pores embedded within an electrically resistant synthetic membrane. This innovative approach has introduced unprecedented opportunities for comprehensive genome analysis, particularly in diseases characterized by complex genomic architecture, structural variation, and extensive genetic heterogeneity, including sickle cell disease (SCD).

The fundamental principle of Oxford Nano pore Sequencing is based on Nano pore electrophysiology. During sequencing, a voltage is applied across a membrane containing Nano scale protein pores, generating an ionic current through each pore. Individual DNA or

RNA molecules are guided through the Nano pore by a processed motor protein that controls their translocation speed. As successive groups of nucleotides occupy the Nano pore, each combination produces a characteristic disruption in the ionic current. These electrical signal fluctuations are continuously recorded and interpreted by sophisticated computational algorithms that convert the raw signal into nucleotide sequences through a process known as base-calling. Unlike conventional sequencing methods that infer DNA sequences indirectly through chemical reactions and imaging, Nano pore sequencing directly measures the physical properties of native nucleic acid molecules, allowing continuous real-time sequencing of exceptionally long DNA fragments.

One of the defining advantages of Oxford Nano pore Sequencing is its capacity to generate long sequencing reads that typically range from several kilo bases to hundreds of kilo bases, with ultra-long reads exceeding one mega base reported under optimized experimental conditions. These extended read lengths enable sequencing across repetitive elements, highly homologous genomic regions, segmental duplications, GC-rich sequences, and structurally complex loci that are often fragmented or incompletely resolved by short-read sequencing platforms. Consequently, long-read sequencing facilitates more accurate genome assembly, haplotype reconstruction, structural variant detection, and characterization of chromosomal rearrangements, making it particularly valuable for investigating inherited disorders with complex genomic landscapes.

Continuous technological improvements have substantially enhanced the performance and clinical utility of Oxford Nano pore sequencing. Advances in Nano pore chemistry, protein pore engineering, sequencing flow cells, library preparation protocols, and neural network–based base-calling algorithms have significantly improved raw-read accuracy while maintaining the unique advantages of long-read sequencing. The introduction of advanced computational models employing deep learning and artificial intelligence has reduced sequencing errors, improved photopolymer resolution, and increased confidence in variant detection. Simultaneously, the development of more efficient bioinformatics workflows has facilitated accurate identification of single nucleotide variants (SNVs), insertions and deletions (indels), structural variants (SVs), copy number variations (CNVs), repeat expansions, and epigenetic modifications from Nano pore sequencing data, thereby expanding its applicability in both research and clinical genomics.

A distinctive feature that differentiates Oxford Nano pore sequencing from most conventional sequencing technologies is its ability to directly sequence native DNA molecules while simultaneously detecting epigenetic modifications. Because DNA methylation and certain other base modifications alter the electrical current generated during Nano pore translocation, specialized signal-processing algorithms can identify these modifications without the need for bisulfide conversion, enzymatic treatment, or additional laboratory procedures. This capability allows simultaneous acquisition of genomic and epigenetic information from the same sequencing experiment, providing valuable insights into gene regulation, chromatin organization, elytroid differentiation, and disease-associated epigenetic mechanisms. The integration of sequence and methylation data has become increasingly important for understanding the molecular basis of phenotypic variability in inherited diseases, including SCD.

The unique characteristics of Oxford Nano pore Sequencing offer several important advantages for investigating the molecular genetics of sickle cell disease. First, long sequencing reads enable accurate phasing of HBB haplotypes, allowing pathogenic variants and modifier alleles to be assigned to specific parental chromosomes. Haplotype phasing is particularly valuable for understanding inheritance patterns, genotype–phenotype relationships, and the interactions between the primary HBB mutation and neighbouring regulatory elements. Second, Nano pore sequencing facilitates the detection of complex structural variants, including inversions, large insertions, deletions, duplications, and translocations within the β-globin gene cluster, which are often difficult to identify using short-read sequencing technologies.

Third, Oxford Nano pore Sequencing substantially improves the identification of rare pathogenic variants located within repetitive, GC-rich, or structurally complex genomic regions that are poorly covered by conventional sequencing platforms. The ability to generate continuous long reads reduces mapping ambiguity and enhances variant-calling accuracy across regions of high sequence homology. Fourth, the technology enables precise characterization of copy number variations (CNVs) and large genomic rearrangements that may influence gene dosage, regulatory interactions, or disease-modifying pathways relevant to SCD. Fifth, direct DNA methylation profiling without bisulfide conversion provides important insights into epigenetic regulation of genes involved in fatal haemoglobin expression, inflammation, oxidative stress, and vascular biology. Finally, the exceptional read length of Nano pore sequencing allows resolution of repetitive sequences and highly homologous genomic regions that remain inaccessible or incompletely assembled using short-read sequencing methods, thereby providing a more complete representation of genomic architecture. Beyond its analytical capabilities, Oxford Nano pore Sequencing offers several practical advantages that support its increasing adoption in clinical genomics. Portable sequencing devices, such as the Minion platform, require relatively modest laboratory infrastructure, produce sequencing data in real time, and permit rapid analysis in a wide range of settings, including hospital laboratories, remote field locations, and resource-limited environments. These characteristics are particularly attractive for regions with a high burden of sickle cell disease, where access to large-scale sequencing facilities may be limited. The portability, scalability, and decreasing cost of Nano pore sequencing have created new opportunities for expanding genomic medicine, new-born screening, carrier testing, and precision diagnostics in underserved populations.

Increasingly, Oxford Nano pore Sequencing is being incorporated into the diagnosis of rare genetic disorders because long-read sequencing consistently improves the detection of disease-causing variants that are frequently missed by conventional molecular approaches. Numerous studies have demonstrated its value in identifying structural variants, repeat expansions, complex genomic rearrangements, and pathogenic haplotypes responsible for inherited neurological, neuromuscular, cardiovascular, haematological, and developmental disorders. In sickle cell disease, these capabilities are particularly important for comprehensive characterization of the β-globin locus, identification of rare disease-modifying variants, pharmacogenomics profiling, and selection of patients who may benefit from emerging gene-editing or gene-replacement therapies. As sequencing chemistry, analytical software, and clinical interpretation frameworks continue to evolve, Oxford Nano pore Sequencing is expected to become an increasingly integral component of precision medicine and personalized healthcare for individuals with SCD and other inherited disorders

V. CLINICAL RELEVANCE OF NANOPORE SEQUENCING IN SICKLE CELL DISEASE

The emergence of Oxford Nano pore Sequencing (ONS) has introduced new opportunities for improving the diagnosis, risk assessment, and clinical management of sickle cell disease (SCD). Although conventional molecular testing has successfully identified the primary pathogenic HBB mutation responsible for SCD, it does not fully explain the extensive variability observed in disease severity, complication profiles, and therapeutic responses among affected individuals. The increasing recognition that SCD is influenced by a complex interaction between the causal HBB mutation, genetic modifiers, structural genomic alterations, epigenetic regulation, and environmental factors has created a need for comprehensive genomic approaches capable of capturing the complete molecular landscape of each patient. By combining long-read sequencing, structural variant detection, haplotype resolution, and epigenetic profiling within a single platform, Oxford Nano pore Sequencing provides a powerful strategy for advancing precision medicine in SCD.

One of the most important clinical applications of Oxford Nano pore Sequencing is the comprehensive identification of pathogenic variants beyond the canonical HBB mutation. Traditional diagnostic approaches primarily focus on detecting the sickle cell mutation within the HBB gene, which is sufficient for confirming diagnosis but insufficient for predicting disease trajectory or explaining phenotypic variability. Long-read sequencing enables simultaneous assessment of coding variants, non-coding regulatory mutations, structural alterations, and other genomic features that may influence disease biology. This comprehensive approach allows clinicians and researchers to identify additional pathogenic or modifying variants that contribute to differences in haemolysis severity, inflammatory responses, vascular complications, and organ damage. Such information may improve molecular classification of patients and support more individualized approaches to disease monitoring and treatment selection.

A second major application involves the improved characterization of modifier genes associated with disease severity. Several common genetic modifiers, including BCL11A, HBS1L-MYB, and KLF1, regulate fatal haemoglobin production and influence clinical outcomes in SCD. However, these known modifiers account for only a portion of phenotypic variation, and many additional rare variants affecting erythropoiesis, inflammation, oxidative stress, endothelial biology, immune regulation, and coagulation remain incompletely characterized. Oxford Nano pore Sequencing enables high-resolution analysis of these genomic regions by providing long reads that preserve regulatory information and facilitate accurate assignment of variants to individual chromosomes. This capability is particularly valuable for understanding how combinations of rare and common variants interact to determine disease severity and treatment response.

Another important clinical contribution of Nano pore sequencing is the detection and characterization of structural variants affecting globin gene regulation. The β-globin gene cluster contains complex regulatory elements, repetitive sequences, and regions of high sequence similarity that are difficult to analyse using short-read sequencing technologies. Structural variants within this region may alter globin gene expression, modify fatal haemoglobin production, or influence the balance between normal and abnormal haemoglobin synthesis. Long-read sequencing provides direct visualization of large genomic rearrangements, including deletions, duplications, inversions, and complex rearrangements involving the β-globin locus. Accurate detection of these variants may improve diagnosis of

atypical haemoglobin disorders, clarify complex genotypes, and enhance interpretation of disease severity in patients with unusual clinical presentations.

Oxford Nano pore sequencing also provides significant advantages for long-range heliotyping and genetic counselling. Because long reads span large genomic regions, they enable direct phasing of variants onto maternal and paternal chromosomes, allowing accurate reconstruction of inherited haplotypes. In SCD, haplotype information can provide valuable insight into the origin and evolution of sickle cell alleles, interactions between disease-causing mutations and modifier variants, and inheritance patterns within families. This information is particularly useful for reproductive counselling, carrier assessment, prenatal diagnosis, and evaluation of recurrence risks among family members. Long-range heliotyping may also improve interpretation of genotype–phenotype relationships by determining whether potentially important modifier variants occur on the same chromosome as the pathogenic HBB allele or on the alternative chromosome.

The expanding use of gene therapy and genome-editing approaches in SCD has created additional clinical applications for long-read sequencing. Recent therapeutic strategies, including CRISPR-based editing and other genetic modification approaches, aim to reactivate fatal haemoglobin production, correct the pathogenic HBB mutation, or introduce protective genetic changes into hematopoietic stem cells. Successful clinical implementation of these therapies requires highly accurate methods for evaluating editing efficiency, confirming intended genomic modifications, detecting unintended structural changes, and monitoring long-term genomic stability. Oxford Nano pore Sequencing offers several advantages in this context because it can analyse large genomic regions surrounding editing sites, identify unexpected insertions or deletions, detect chromosomal rearrangements, and evaluate potential off-target effects that may be missed by targeted short-read sequencing methods.

Another emerging application is the simultaneous assessment of sequence variation and DNA methylation patterns. Epigenetic regulation plays a critical role in elytroid development, globin gene switching, inflammatory activation, and cellular stress responses. Conventional sequencing methods generally require separate laboratory procedures to measure genetic variation and DNA methylation, increasing cost and complexity. In contrast, Oxford Nano pore Sequencing can directly detect methylated bases from native DNA molecules during sequencing, allowing integrated analysis of genetic and epigenetic information. In SCD, this capability may improve understanding of regulatory mechanisms controlling fetal haemoglobin expression, inflammatory pathways, and treatment response. Such combined genomic and epigenetic profiling may ultimately contribute to improved patient stratification and identification of novel therapeutic targets.

Nano pore sequencing also has considerable potential for enhancing molecular diagnosis in genetically unresolved cases. A proportion of patients with suspected inherited blood disorders remain without definitive molecular explanations after conventional genetic testing due to limitations in detecting structural variants, repetitive regions, complex rearrangements, or regulatory mutations. Long-read sequencing can overcome these limitations by providing a more complete representation of genomic architecture and enabling discovery of previously undetected disease-associated variants. This is particularly relevant in SCD patients with atypical clinical presentations, unusual haemoglobin profiles, unexpected treatment responses, or discordance between genotype and phenotype.

Improved molecular resolution may facilitate more accurate diagnosis, reduce uncertainty, and guide personalized clinical management.

Beyond diagnostic applications, Oxford Nano pore Sequencing has important implications for future disease monitoring and personalized treatment strategies. As genomic medicine moves toward individualized care, comprehensive molecular profiling may allow clinicians to identify patients at increased risk of severe complications, optimize therapeutic interventions, and monitor disease progression more accurately. Integration of nanopore sequencing with electronic health records, artificial intelligence-based variant interpretation, and multi-omits approaches may further enhance predictive models of disease outcomes. In regions with high SCD prevalence and limited access to advanced genomic infrastructure, portable Nano pore platforms also offer opportunities for decentralized testing and implementation of precision medicine approaches closer to affected communities.

As gene therapy and genome-editing technologies become increasingly integrated into SCD management, Oxford Nano pore Sequencing is expected to play an essential role in ensuring therapeutic safety, accuracy, and long-term effectiveness. Its ability to evaluate genome-wide structural integrity, detect off-target events, confirm editing outcomes, and monitor genomic stability over time makes it a valuable component of next-generation therapeutic development. Although challenges remain regarding cost, standardization, regulatory approval, sequencing accuracy, and clinical interpretation, continued technological improvements are likely to establish Nano pore sequencing as a central tool in the transition from traditional disease management toward comprehensive precision medicine for sickle cell disease.

VI. CONCLUSION

The future of sickle cell disease (SCD) research and clinical management is increasingly moving beyond the traditional concept of a single-gene disorder toward a more comprehensive understanding of the broader genomic, epigenetic, and biological landscape that determines individual disease expression. Although the pathogenic HBB Glu6Val mutation remains the fundamental molecular cause of SCD, decades of clinical observations have demonstrated that this mutation alone cannot explain the remarkable diversity in disease severity, complication patterns, treatment responses, and survival outcomes observed among affected individuals. The integration of genomic studies has revealed that common genetic modifiers, rare pathogenic variants, structural genomic alterations, regulatory elements, and epigenetic mechanisms collectively contribute to the complex phenotype of SCD. This evolving perspective emphasizes the need for comprehensive genomic characterization to achieve accurate risk prediction and truly personalized disease management.

Oxford Nano pore Sequencing (ONS) has emerged as one of the most powerful technologies for addressing these challenges because of its unique ability to analyse long native DNA molecules and capture multiple layers of genomic information within a single sequencing platform. Unlike conventional short-read sequencing approaches, Nano pore sequencing enables simultaneous detection of single nucleotide variants, insertions and deletions, structural rearrangements, copy number variations, repeat expansions, long-range haplotypes, and DNA methylation patterns. This comprehensive analytical capacity provides an unprecedented opportunity to resolve complex genomic regions, including the β-globin

gene cluster, and to identify rare pathogenic variants that may contribute to disease severity but remain undetected using traditional molecular approaches.

The application of Oxford Nano pore sequencing in sickle cell disease has the potential to transform molecular diagnosis and patient stratification by providing a more complete representation of each individual's genomic profile. Identification of rare modifier variants and structural alterations may improve prediction of clinical complications, including vase-occlusive crises, stroke, pulmonary hypertension, renal disease, and progressive organ damage. Furthermore, integration of genomic and epigenetic information may enhance understanding of regulatory mechanisms controlling fatal haemoglobin expression, inflammatory pathways, erythrocyte biology, and treatment responsiveness. Such knowledge could support the development of personalized therapeutic strategies, allowing clinicians to move beyond generalized treatment approaches toward interventions tailored to each patient's unique molecular characteristics.

The expanding role of Nano pore sequencing is particularly relevant as advanced therapeutic approaches, including hematopoietic stem cell transplantation, gene addition therapies, and CRISPR-based genome-editing strategies, become increasingly available for patients with SCD. These emerging treatments require highly accurate genomic monitoring to confirm therapeutic modifications, evaluate editing efficiency, identify unintended genomic alterations, and ensure long-term safety. Oxford Nano pore Sequencing provides an important advantage in this context because its long-read capability enables comprehensive evaluation of edited genomic regions, detection of large-scale rearrangements, and monitoring of genomic stability over time. As gene-based therapies continue to evolve, long-read sequencing technologies are likely to become essential components of therapeutic development, regulatory evaluation, and clinical follow-up.

Despite its considerable promise, several challenges must be addressed before Oxford Nano pore Sequencing becomes routinely integrated into clinical practice for SCD. These challenges include continued improvement of sequencing accuracy, development of standardized laboratory protocols, optimization of bioinformatics pipelines, establishment of clinically validated variant interpretation frameworks, reduction of costs, and expansion of population-specific genomic reference databases. These issues are particularly important for populations most affected by SCD, including individuals of African, Middle Eastern, and South Asian ancestry, because existing genomic resources remain disproportionately derived from European populations. Ensuring equitable access to advanced genomic technologies and incorporating diverse populations into genomic research will be essential for maximizing the clinical benefits of precision medicine.

Ethical considerations will also play a central role in the implementation of Nano pore-based genomic medicine. Comprehensive sequencing generates large amounts of sensitive genetic information, raising important questions regarding data privacy, informed consent, ownership of genomic data, interpretation of incidental findings, and equitable distribution of healthcare benefits. Responsible integration of long-read sequencing into SCD management will require collaboration among researchers, clinicians, patients, policymakers, and healthcare systems to establish ethical frameworks that protect individuals while promoting scientific advancement.

Looking forward, continued improvements in Nano pore sequencing chemistry, artificial intelligence-based base-calling, automated variant interpretation, multi-omics integration,

and cloud-based genomic analysis are expected to further enhance the clinical utility of this technology. The combination of genomic, epigenetic, transcript omic, and clinical data may enable the development of predictive models capable of identifying patients at high risk for complications and guiding early intervention strategies. In resource-limited regions where SCD represents a major health burden, portable Nano pore sequencing platforms may also provide opportunities for decentralized genomic testing, improved new-born screening, and expanded access to precision healthcare.

In conclusion, Oxford Nano pore Sequencing represents a major advancement in the genomic investigation of sickle cell disease by enabling detailed characterization of the complex molecular factors that influence individual disease outcomes. Its ability to uncover rare pathogenic variants, resolve structural genomic complexity, and integrate epigenetic information provides a foundation for the next generation of precision medicine approaches. As technological and analytical barriers continue to decrease, Nano pore sequencing is expected to become an increasingly important tool for improving diagnosis, prognosis, and therapeutic decision-making, and long-term management of individuals living with sickle cell disease worldwide.

REFERENCES

[1] Amarasinghe, S. L., Su, S., Dong, X., Zappia, L., Ritchie, M. E., & Gouil, Q. (2020). Opportunities and challenges in long-read sequencing data analysis. Genome Biology, 21, Article 30.

[2] Anna, A., & Monika, G. (2018). Splicing mutations in human genetic disorders: Examples, detection, and confirmation. Journal of Applied Genetics, 59(3), 253–268.

[3] Bamshad, M. J., Ng, S. B., Bigham, A. W., Tabor, H. K., Emond, M. J., Nickerson, D. A., & Shendure, J. (2011). Exome sequencing as a tool for Mendelian disease gene discovery. Nature Reviews Genetics, 12(11), 745–755.

[4] Cirulli, E. T., & Goldstein, D. B. (2010). Uncovering the roles of rare variants in common disease through whole-genome sequencing. Nature Reviews Genetics, 11(6), 415–425.

[5] De Coster, W., Weissensteiner, M. H., & Sedlazeck, F. J. (2021). Towards population-scale long-read sequencing. Nature Reviews Genetics, 22(9), 572–587.

[6] Eaton, W. A., & Hofrichter, J. (1990). Sickle cell hemoglobin polymerization. Advances in Protein Chemistry, 40, 63–279.

[7] Frangoul, H., Altshuler, D., Cappellini, M. D., Chen, Y. S., Domm, J., Eustace, B. K., Foell, J., de la Fuente, J., Grupp, S., Handgretinger, R., Ho, T. W., Kattamis, A., Kernytsky, A., Lekstrom-Himes, J., Li, A. M., Locatelli, F., Mapara, M. Y., de Montalembert, M., Rondelli, D., … Corbacioglu, S. (2021). CRISPR-Cas9 gene editing for sickle cell disease and β-thalassemia. New England Journal of Medicine, 384(3), 252–260.

[8] Gladwin, M. T., Sachdev, V., Jison, M. L., Shizukuda, Y., Plehn, J. F., Minter, K., Brown, B., & Ognibene, F. P. (2004). Pulmonary hypertension as a risk factor for death in patients with sickle cell disease. New England Journal of Medicine, 350(9), 886–895.

[9] Goodwin, S., McPherson, J. D., & McCombie, W. R. (2016). Coming of age: Ten years of next-generation sequencing technologies. Nature Reviews Genetics, 17(6), 333–351.

[10] Ingram, V. M. (1957). Gene mutations in human haemoglobin: The chemical difference between normal and sickle cell haemoglobin. Nature, 180, 326–328.

[11] Jain, M., Olsen, H. E., Paten, B., & Akeson, M. (2018). The Oxford Nanopore MinION: Delivery of nanopore sequencing to the genomics community. Genome Biology, 19, Article 239.

[12] Kato, G. J., Piel, F. B., Reid, C. D., Gaston, M. H., Ohene-Frempong, K., Krishnamurti, L., Smith, W. R., Panepinto, J. A., Weatherall, D. J., Costa, F. F., & Vichinsky, E. P. (2018). Sickle cell disease. Nature Reviews Disease Primers, 4, Article 18010.

[13] Lettre, G., Sankaran, V. G., Bezerra, M. A. C., Araújo, A. S., Uda, M., Sanna, S., Cao, A., Schlessinger, D., Costa, F. F., Hirschhorn, J. N., & Orkin, S. H. (2008). DNA polymorphisms at the BCL11A, HBS1L-MYB, and β-globin loci associate with fetal hemoglobin levels and pain crises in sickle cell disease. Proceedings of the National Academy of Sciences, 105(33), 11869–11874.

[14] Liu, Q., Fang, L., Yu, G., Wang, D., Xiao, C. L., & Wang, K. (2019). Detection of DNA base modifications by deep recurrent neural network on Oxford Nanopore sequencing data. Nature Communications, 10, Article 2449.

[15] Logsdon, G. A., Vollger, M. R., & Eichler, E. E. (2020). Long-read human genome sequencing and its applications. Nature Reviews Genetics, 21(10), 597–614.

[16] Menzel, S., Garner, C., Gut, I., Matsuda, F., Yamaguchi, M., Heath, S., Foglio, M., Zelenika, D., Boland, A., Rooks, H., Best, S., Spector, T. D., Farrall, M., Lathrop, M., & Thein, S. L. (2007). A QTL influencing fetal haemoglobin production maps to a gene encoding a zinc-finger protein on chromosome 2p15. Nature Genetics, 39(10), 1197–1199.

[17] Payne, A., Holmes, N., Clarke, T., Munro, R., Debebe, B. J., & Loose, M. (2021). Readfish enables targeted Nano pore sequencing of gig abase-sized genomes. Nature Biotechnology, 39(4), 442–450.

[18] Piel, F. B., Patil, A. P., Howes, R. E., Nyangiri, O. A., Gething, P. W., Dewi, M., Temperley, W. H., Williams, T. N., Weather all, D. J., & Hay, S. I. (2013). Global epidemiology of sickle haemoglobin in neonates: A contemporary geostatistical model-based map and population estimates. The Lancet, 381(9861), 142–151.

[19] Rees, D. C., Williams, T. N., & Gladwin, M. T. (2010). Sickle-cell disease. The Lancet, 376(9757), 2018–2031.

[20] Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., Grody, W. W., Hegde, M., Lyon, E., Spector, E., Voelkerding, K., & Rehm, H. L. (2015). Standards and guidelines for the interpretation of sequence variants: A joint consensus recommendation of the ACMG and AMP. Genetics in Medicine, 17(5), 405–424.

[21] Roadmap Epigenetics Consortium. (2015). Integrative analysis of 111 reference human epigones. Nature, 518(7539), 317–330.

[22] Simpson, J. T., Workman, R. E., Zuzarte, P. C., David, M., Dursi, L. J., & Timp, W. (2017). Detecting DNA cytosine methylation using Nano pore sequencing. Nature Methods, 14(4), 407–410.

[23] Steinberg, M. H., & Sebastiani, P. (2012). Genetic modifiers of sickle cell disease. American Journal of Haematology, 87(8), 795–803.

[24] Sudmant, P. H., Rausch, T., Gardner, E. J., Handsaker, R. E., Abyzov, A., Huddleston, J., Zhang, Y., Ye, K., Jun, G., Fritz, M. H. Y., Konkel, M. K., Malhotra, A., Stütz, A. M., Shi, X., Casale, F. P., Chen, J., Hormozdiari, F., Dayama, G., Chen, K.,1000 Genomes Project Consortium. (2015). an integrated map of structural variation in 2,504 human genomes. Nature, 526(7571), 75–81.

[25] Sundd, P., Gladwin, M. T., & Novelli, E. M. (2019). Pathophysiology of sickle cell disease. Annual Review of Pathology: Mechanisms of Disease, 14, 263–292.

[26] Ware, R. E., de Mont Alembert, M., Tshilolo, L., & Abboud, M. R. (2017). Sickle cell disease. The Lancet, 390(10091), 311–321.

[27] Wick, R. R., Judd, L. M., & Holt, K. E. (2019). Performance of neural network base calling tools for Oxford Nano pore sequencing. Genome Biology, 20, Article 129.