Evaluation of low-cost techniques for detecting sickle cell disease and β-thalassemia in Nepal

1.0 Background

1.1 Sickle cell disease

Sickle cell disease (SCD) is a collection of inherited blood disorders, and is a multisystem disorder affecting nearly every organ in the body [1]. The disease primarily causes erythrocytes or red blood cells (RBCs) to be distorted and rigid due to polymerization (e.g. during deoxygenation), often changing them to crescent or sickle shaped cells [2]. Other cascading effects include hemolytic anemia, blockage of microvasculature (capillaries and small blood vessels), episodes of acute illness (such as vaso-occlusive pain crisis or sickle cell crisis), and progressive organ damage [3], [4]. The chances of survival and quality of life drastically increase with early diagnosis, newborn screening, penicillin prophylaxis, pneumococcal immunization, blood transfusion, hydroxyurea treatment, hematopoietic stem-cell transplantation, and education about disease severity and complications [1], [4].

SCD is one of the most common monogenic diseases worldwide and results from mutations in the β-globin gene [1], [3]. The most common and severe form of SCD is sickle cell anemia (SCA), caused by the inheritance of β^S alleles or the variant HbS (hemoglobin S or sickle hemoglobin) genes from both parents – resulting in the homozygous form HbSS, instead of the normal homozygous form HbAA. Another form of SCA is HbS/β^0 thalassemia, caused by the inheritance of a β^0 thalassemia mutation from one parent and an HbS gene from another, resulting in no normal HbA and a majority of HbS. The heterozygous form HbAS, known as sickle cell trait (SCT), affects millions of people worldwide and is usually not associated with hematological abnormalities [2]. On the other hand, compound heterozygous conditions are associated with mild to severe clinical courses and are also categorized as SCD. Some compound heterozygous conditions include HbS with hemoglobin C or HbC (HbSC), HbS with β-thalassemia (HbS/β^0-thalassemia or HbS/β^+-thalassemia), and HbS with other β globin variants (*e.g.* HbSD and $HbSO^{Arab}$) [1], [3], [4].

An increasing global health problem, SCD is associated with an estimated 300,000 to 400,000 births every year [5]. The prevalence of the disease is high in sub-Saharan Africa, the Middle East, the Mediterranean basin and the Indian subcontinent, mainly due to the remarkable survival advantage of the sickle cell trait against severe malaria ("malaria hypothesis") [1]. It is estimated that people with SCT (heterozygous HbAS) are 90% less likely to experience severe malaria than individuals with normal hemoglobin [5] – making the malaria hypothesis a classic example of natural selection and balanced polymorphism [1]. The sickle hemoglobin allele (HbS) has spread beyond malaria-endemic regions due to population movements (including slave trading), especially to regions in North America and Western Europe [5]. Additionally, gene frequencies of α and β thalassemia can be greater than 80% in malaria-endemic regions [2].

1.2 Thalassemia and its interaction with sickle cell disease

Hemoglobin is the protein in red blood cells that transports oxygen (and carbon dioxide) between lungs and other tissues/cells in the body. The hemoglobin molecule is made up 4 polypeptide chains or subunits, such as alpha, beta, gamma or delta chains. The three most common types of hemoglobin are hemoglobin A (normal adult hemoglobin), hemoglobin F (fetal hemoglobin), and hemoglobin A2 (normally found in small quantities in adults). Hemoglobin A is

made up of 2 alpha chains and 2 beta chains ($\alpha_2\beta_2$). Hemoglobin F is made up of 2 alpha chains and 2 gamma chains ($\alpha_2\gamma_2$). Hemoglobin A2 is made up of 2 alpha chains and 2 delta chains ($\alpha_2\delta_2$).

Thalassemia is an inherited condition which results in a deficiency or abnormality in the synthesis of one or more of these polypeptide/globin chains or subunits. Clinical symptoms include anemia resulting from inadequate hemoglobin synthesis, and ineffective production of red blood cells due to an unbalanced accumulation of one of the globin chains. α-thalassemia is caused by mutations in the alpha globin genes. Since each chromosome normally has 2 alpha genes, there are four possible syndromes related to α-thalassemia with increasing severity caused due to 1 to 4 genes being absent or failing to function: 1) α^+ thalassemia trait, 2) α^0 thalassemia trait, 3) hemoglobin H disease, and 4) hemoglobin Bart's hydrops fetalis. β-thalassemia is caused by mutations in the beta globin genes. Although there are more than 400 mutations identified, the majority of cases either result in a reduced (β^+) or absent (β^0) production of beta chains. Based on the clinical severity, thalassemia is also classified as thalassemia minor or trait (where individuals are silent carriers and asymptomatic), thalassemia intermedia (associated with mild anemia and is transfusion independent), and thalassemia major (associated with anemia and is transfusion dependent) [2].

The interaction and clinical severity of sickle cell with β-thalassemia depends on the type of β-thalassemia gene inherited. Coinheritance of HbS and β^+-thalassemia results in mild sickling disorder, while coinheritance of HbS and β^0-thalassemia results in severe sickle cell disease. The clinical symptoms and test results of HbS/β^0-thalassemia is similar to HbSS (homozygous SCD) because both cases predominantly have HbS and no HbA, which makes it difficult to distinguish the two cases. The clinical syndromes and levels of anemia related to HbSS and α-thalassemia are generally less severe than those related to HbS/β^0-thalassemia. Generally, a distinction between HbSS with and without α-thalassemia is clinically not required. The interaction of HbS with other hemoglobin variants such as HbD^{Punjab} or HbO^{Arab} also results in severe sickle cell disease [2].

1.3 Global burden of disease

Over the next 30 years, it is estimated that 15 million babies will be born with SCD globally, and 50-90% of infants born with SCD in low-income countries will die before 5 years of age [6]. The burden of the disease is much greater in lower-resource settings [4], with around 90% of SCD occurring in low- and middle-income countries [7]. The highest prevalence has been found in sub-Saharan Africa, where approximately 230,000 children are born with SCD every year (0.74% of births in the region). In contrast, 2600 and 1300 children are born with SCD in North America and Europe, respectively [3]. In addition to screening for SCD, it is also essential to screen for SCT (heterozygous HbAS) because parents with SCT can give birth to children with SCD (homozygous HbSS or compound heterozygous forms). The number of babies born with SCT worldwide is estimated to be approximately 5.5 million, and the prevalence is estimated to be 2-30% in African countries [4]. Additionally, around 130,000 children are born with severe hemoglobinopathies such as HbSC and thalassemia annually [8].

In countries with newborn screening and treatment options, like the US and UK, the survival of children with SCA is similar to that of healthy children and adults can live up to their sixties [5]. However, in low-resource settings without newborn screening and treatments such as penicillin prophylaxis and pneumococcal immunization, the mortality rate for children with SCD is tremendously high [4]. In addition, the mortality rates are inaccurate due to lack of adequate screening, and improper diagnosis and management – SCD can be a predisposing factor for

deaths resulting from pneumonia, malaria, tuberculosis, HIV, and diarrheal disease [4], [8]. SCD does not just adversely affect the people involved, but also their families, with increased financial and psychosocial burden [8].

The number of people with SCD is expected to increase worldwide and by 2050, the number of babies affected is estimated to increase by 30%, thus increasing the health burden of the disease [4], [5]. In high income countries, the increase can be attributed to increase in life expectancy due to newborn screening, hydroxyurea treatment, and penicillin prophylaxis. In low-income and middle-income countries, this increase is due to epidemiological shifts such as public health improvements, and changes in population distribution, mortality, fertility, life expectancy, and infant mortality [1], [5]. As the burden of the disease increases, early detection of the disease is a critical first step towards managing the disease, especially in lower-income countries.

1.4 Need for screening/diagnosis in remote and rural settings

One of the first steps to addressing SCD in low-income countries is to break the diagnostic barrier. Since current laboratory techniques are too expensive to enable timely and equitable diagnosis, low-cost and rapid point-of-care diagnostic devices are required for early detection [9]. There have been promising pilot programs for neonatal screening in many sub-Saharan countries, however, few have translated to sustainable national programs. Newborn screening has shown high cost-effectiveness in Angola, and there is also compelling evidence of reduced infant mortality. However, most SCD detection and management programs are located in urban areas, and not in rural areas, where the prevalence is higher (as observed in rural populations of Africa and marginalized tribal populations in India) [4]. Thus, cost-effective screening and management services need to be prioritized in rural and remote settings to address the disparity.

Government support is vital to help patients detected with the disease in low-income countries. The cost of one SCD-related hospitalization exceeds the average monthly income of a household in African countries [8]. A one-month supply for hydroxyurea treatment in Ghana costs around USD $100, which the government recently planned to subsidize as part of the Ghana's national health insurance scheme [6]. Even the Nepalese government has provisioned NRs 1,00,000 (Nepalese Rupees) for health care related expenses to patients identified with SCA [10]. However, with limited screening programs in rural/remote areas, most of the affected patients do not have access to such initiatives, yet. For maximizing impact of current government initiatives and for encouraging new sustainable government programs, low-cost and accurate SCD screening and diagnostic techniques must be implemented in rural and remote settings.

In areas where HbS is prevalent with other hemoglobin variants (such as β-thalassemia, HbC, etc.), it is essential to identify all the hemoglobin variants possible. If screening techniques that only detect HbS are used in such areas, then compound heterozygous cases (such as HbS/β-thalassemia) could be wrongly diagnosed. Additionally, asymptomatic heterozygous cases (such as HbA/β-thalassemia) may be missed in the screening tests for HbS, and parents with such heterozygous cases can still possibly pass on the variant gene to their children to result in compound heterozygous forms of SCD. Recent studies have indicated a high prevalence of β-thalassemia in Nepal, with approximately 26% - 70% of hemoglobinopathies involving β-thalassemia [11], [12]. Therefore, screening techniques in Nepal should ideally include detection of β-thalassemia in addition to SCD/SCT.

1.5 Conventional screening and diagnosis

The following techniques have been commonly used to detect hemoglobin S (HbS) related to SCD or SCT.

Sickling test: The presence of hemoglobin S can be demonstrated in a sickling test, where red blood cells change their shape from round bi-concave shape to sickled shape, under conditions of hypoxia or deoxygenation. When a drop of blood is sealed (using petroleum jelly/paraffin wax mixture or nail lacquer) between a glass slide and a cover slip, the metabolic activity of white blood cells causes deoxygenation, which induces sickling after a few hours, if HbS is present. The deoxygenation can be accelerated by adding a reducing agent such as sodium metabisulphite or sodium dithionite. Sickling can be induced using a reducing agent within 30 minutes to an hour for SCD and SCT, and the changes in morphology of red blood cells can be observed under a microscope, to distinguish SCD and SCT individuals from healthy participants [2], [13]. The sensitivity of the sickling test ranges from 65% [14] to 97.3% [15] and the specificity ranges from 96% [14] to 99.6% [15]. The sickling test, which is relatively inexpensive, was predominantly used as a screening test for HbS, but has mostly been replaced by the solubility test.

HbS solubility test: The solubility test relies on the principle that hemoglobin S, when deoxygenated, is insoluble in a high molarity phosphate buffer, which forms crystals that refract light and causes the solution to turn turbid [2]. Sickle solubility tests provide positive results for cases of SCD and SCT beyond early infancy – for early infancy (infants less than 6 months) the percentage of HbS can be below 20%, which could result in false negative tests [13]. The sickle solubility test, whether positive, negative or ambiguous, should be followed up by a confirmatory test, such as hemoglobin electrophoresis. The follow up tests are required to confirm the presence of HbS, distinguish SCT from SCD, or identify false negative cases due to low HbS concentration or technical error [13]. For rapid results or for emergency cases, the sickle solubility test can be combined with a blood count and a blood film analysis to distinguish SCT from SCD with reasonable accuracy. The sensitivity of the solubility test ranges from 45% [14] to 99% [15] and the specificity ranges from 90% [14] to 99.9% [15].

Hemoglobin electrophoresis (HE): HE relies on the principle that proteins can separate when placed in a charge gradient and can be visually differentiated using protein or heme stains. Several media can be used for performing HE to detect variant hemoglobin, such as cellulose acetate membrane, filter paper, starch gel, citrate agar gel or agarose gel. In the past, hemoglobin electrophoresis was the most common method to initially detect variant hemoglobins, but is currently replaced by high performance liquid chromatography (HPLC) and capillary electrophoresis (CE). Nevertheless, the technique still serves as an essential confirmatory test [13].

Capillary electrophoresis (CE): CE separates normal and variant hemoglobin in alkaline buffers inside capillary tubes. Due to the use of capillary tubes and slightly different pH, the voltage used is higher and the run-time is shorter than other HE techniques. The zone or position of hemoglobin variants with respect to hemoglobin A is used to identify types of hemoglobin. CE only uses small volumes of blood, thus is beneficial for neonatal screening programs [2], [13].

High performance liquid chromatography (HPLC): HPLC is currently a common initial diagnostic technique for hemoglobinopathies in laboratories of high-income countries. The process used in HPLC begins with a mixture of hemoglobin molecules with a net positive charge

adsorbed onto a negatively charged stationary region. This is followed by elution of the molecules by a liquid with increasing amounts of cations, which compete with the hemoglobin molecules. As the adsorbed hemoglobin molecules get eluted, they get optically detected. The type of hemoglobin is identified by the characteristic amount of time (retention time) taken for the hemoglobin molecules to be eluted [13]. Although HPLC equipment and reagents cost more than other electrophoresis techniques, the labor costs are reduced due to increased automation, thus resulting in similar overall costs. HPLC also utilizes small sample volumes (around 5 µl), making it useful for neonatal screening [2].

1.6 Recent advancements and novel technologies

Several approaches are currently being developed for low-cost point-of-care screening of SCD and SCT. Some of the most promising technologies include lateral flow immunoassays, portable electrophoresis machines, smartphone-based imaging, low-cost microscopy, and automated detection using machine learning.

Lateral flow immunoassays

HemoTypeSC and Sickle SCAN are promising lateral-flow assays developed in the last few years, with high overall diagnostic accuracies of around 99%, using HPLC or HE as reference tests. Both the lateral-flow assays detect hemoglobin A, S, and C, and are thus able to detect the compound heterozygous case of HbSC. Most of the laboratory and field studies report sensitivities and specificities of 92% - 99% for detecting hemoglobin variants (HbA, HbS, HbC) or phenotypes (HbAA, HbAS, HbSS, HbAC, HbSC, HbCC) [16]–[22]. In a recent study implementing newborn screening in Nigeria, HemoTypeSC, SickleSCAN and HPLC showed concordance (100% sensitivities and specificities) [23]. In remote and rural settings, both assays have been found to be more feasible to implement than HPLC or electrophoresis tests.

HemoTypeSC is a competitive lateral flow immunoassay, costs around USD $2 per test, and has a turnaround time of 10 minutes for results. Each kit contains 50 test strips, and once the kit is opened, the test strips are stable for 30 days. Sickle SCAN is a qualitative lateral flow immunoassay (thus easier to interpret), costs around USD $5, and has a turnaround time of 5 minutes for results. Sickle SCAN kits are provided as individually packaged cartridges, with an unopened shelf life of around 2 years.

Smartphone microscopy and low-cost microscopy

With the ubiquity of smartphones, multiple point-of-care technologies have been developed that utilize their capabilities, such as using their cameras for imaging and disease detection. By attaching lenses externally, the optical capabilities can be enhanced to image microscopic objects. Smartphone microscopy has been implemented to detect SCD by a few groups. de Haan *et al.* identified sickle cells in blood smears of individuals with SCD using a smartphone microscope and two neural networks, with an accuracy of around 98% [24].

In addition to imaging peripheral blood smears, smartphone microscopes have been used to image blood samples mixed with sodium metabisulphite to induce sickling in hypoxic conditions, as is done in a sickling test. Ilyas et al. observed similar percentage of sickle cells with a smartphone-based technique and a conventional microscope, when 1 µL of blood, diluted in phosphate buffered saline (1:100 ratio), was mixed with 2% sodium metabisulphite in a microfluidic device [25]. D'Costa *et al.* used a similar approach of smartphone-based imaging in a microfluidic chamber, but instead used two different concentrations of sodium metabisulphite on blood diluted in a cell culture media. Low concentrations of the reducing agent induced sickling only in SCD blood, while high concentrations induced sickling in SCD and SCT blood

(and not normal blood) [26]. Unlike traditional sickling test, which does not distinguish SCT and SCD patients, this modification allowed them to distinguish SCT, SCD and healthy individuals using simple image analysis techniques. Additionally, the characteristic shapes of the sickled cells differed between individuals with SCT and SCD, indicating that the morphology differences during a sickling test could potentially be used distinguish SCT and SCD patients.

Other inexpensive microscopes compatible with smartphone imaging have also been developed such as Foldscope [27], which is an origami-based paper microscope costing around $2, and CellScope [28], which is created by attaching a reversed camera lens onto a mobile phone camera. Additionally, microscopes with higher resolution lenses, costing between $300 - $600, have been developed using inexpensive parts [29], opto-mechanical components [30], 3D printed parts or common supplies from hardware stores [31]. Some groups have also focused on open sourced designs, such as a modular open source microscopy platform [32], and a one-piece 3D-printed translation stage for microscopy [33].

Gazelle Diagnostic Device

A paper-based microchip electrophoresis platform, called Hemechip, has also been developed to identify hemoglobin S, C, and E variants with 100% sensitivity and overall diagnostic accuracy of 98.4% [34]. The electrophoresis test or the “Gazelle Hb Variant test” is performed in a portable reader called Gazelle Diagnostic Device, developed by Hemex Health (Portland, OR, US). The portable reader, which also enables detection of malaria, includes an imaging system, touch screen device and a rechargeable battery [35]. The cost of the portable reader is USD $1200, and each test (or microchip) costs around USD $2. The Gazelle Hb Variant test detects sickle hemoglobin and β-thalassemia.

Machine Learning

Numerous algorithms have been developed to detect SCD using images of peripheral blood smears [24], [36]–[39]. Early algorithms used image processing methods such as thresholding and clustering to categorize cells, while recent works primarily utilize machine learning techniques for increased accuracy. Specifically, convolutional neural networks can classify images with high accuracy, as long as the sample size is sufficient. Convolutional neural networks are complex algorithms that are “taught” to perform a certain task, such as classifying an image, by using training data to optimize parameters. The typical approach for SCD detection is for the neural network to first segment each individual cell in the image. The cells are then classified into predefined categories and the algorithm determines if sufficient sickle cells are present for a positive screen result.

In one instance [24], an additional neural network was used to enhance images from a low-cost smartphone-based microscope. The research team imaged the same section of a slide with both the smartphone-based microscope and a lab-grade benchtop microscope, then used the pairs of images to train a neural network. This helped to both improve image quality and standardize images, as there were variations among the images taken on the smartphone.

Overall, the existing machine learning approaches require a high-quality peripheral blood smear. Overlapping cells are challenging to segment and classify (although certain algorithms attempt to account for this), meaning a near-perfect monolayer must be present. This presents a limitation for remote settings, as manually creating consistent blood smears requires considerable skill. Additionally, individuals with SCT, or even SCD, often have very few sickle cells in fresh whole blood. Therefore, even an algorithm that can match or exceed the performance of a hematopathologist will have a considerable number of false-negatives.

Currently, machine learning has been implemented in peripheral blood smears but not on deoxygenated blood films prepared during sickling tests, where much more sickle cells appear on both SCT and SCD blood samples than on a blood smear. The conventional sickling test can classify SCT/SCD patients from healthy participants, but is not typically used to distinguish between SCT and SCD under manual observation. D'Costa *et al.* [26] recently slightly modified the sickling test by using two different concentrations of sodium metabisulphite with diluted blood on a microfluidic chamber, and were able to distinguish between SCD and SCT using image analysis. They also found that the distribution of morphology of sickle cells under hypoxia was different for blood from SCD and SCT patients. This was attributed to differences in HbS polymerization depending on the concentration of hemoglobin S and the rate of deoxygenation, which has also been observed earlier [40]. The differences in morphology between SCD and SCT can potentially be used to classify SCD, SCT and healthy participants using the sickling test, combined with machine learning, which is one of the aims of this study.

Other approaches
A paper-based test has been implemented as a low-cost screening technique for SCD. The test involves mixing blood with a solubility/lysis buffer, filtering cellular debris, adding a deoxygenation buffer, and depositing a drop on chromatography paper. Differences in HbS concentration leads to identifiable differences in patterns of blood stains, resulting in detection of HbS with accuracies of around 80% [41]. Another technique involves a 3D printed magnetic levitation platform paired with a smartphone to identify SCD and healthy blood based on differences in densities of red blood cells with and without HbS [42].

1.7 Current initiatives in Nepal

The University of British Columbia (UBC) Sickle Cell Project, a part of the UBC Global Health Initiative, aims to spread awareness of sickle cell disease and to help with screening patients in rural communities of Nepal, while collaborating with local non-profit organizations such as Creating Possibilities (CP) Nepal. CP Nepal has operated since 2006, with an overarching goal of protecting and advocating for the rights of marginalized women and children in Nepal. Since 2015, the UBC Sickle Cell Project have collaborated with CP Nepal, and sent teams of 6-12 medical students annually to conduct screening, diagnosis and education of SCD in the district of Dang, Nepal. So far, four sustainable health posts have been set up in rural parts of Western Nepal to screen for SCT and SCD.

In 2015/2016, a study conducted by the UBC Sickle Cell team and CP Nepal found the prevalence of SCD and SCT to be around 10% within the Tharu community in Dang, Nepal, based on screening approximately 3000 individuals aged 6 months to 40 years [43]. In some parts of Western Nepal, such as Nepalgunj, the prevalence was found to be as high as around 25% [11]. Since many of the studies in Nepal do not include newborn screening, the prevalence of the disease could be higher in rural communities of Nepal, and extensive screening and diagnosis programs are required to accurately identify and treat affected patients. The current study to evaluate performance of low-cost detection techniques is inspired by efforts of teams such as UBC Sickle Cell team and CP Nepal. The UBC Sickle Cell team and CP Nepal plan to help and support the Nepal study.

When the UBC Sickle Cell team conducted their initial study in 2015/2016, they focused on the sickling test for screening and had to send blood samples from some of the patients to India for confirmatory electrophoresis tests, since the gold standard HPLC tests were not available anywhere in Nepal. Over the last few years, some hospitals and clinics across Nepal have

added HPLC machines and have introduced tests for hemoglobinopathies. In Nepalgunj, such HPLC machines and tests have been introduced at Bheri Hospital (where diagnosis is currently free of charge for SCD patients) and Mount Sagarmatha polyclinic. Dr. Rajan Pande (Chief Consultant Physician at Bheri Hospital) played an active role in advocating for the importance of testing hemoglobinopathies in Nepal based on his experience in diagnosing SCD among people in and around Nepalgunj. Dr. Rajan Pande will help as a principal investigator in Nepal. The tests in Nepal, including HPLC tests, will be performed at Mount Sagarmatha Polyclinic at Nepalgunj.

1.8 Need/justification for project

Though early SCD and β-thalassemia detection is a prerequisite for drastically improving patient outcomes, rural and remote communities in low- and middle-income countries often lack effective screening and diagnostic programs. There is a need to evaluate the feasibility of point-of-care techniques to detect SCD, SCT and β-thalassemia. Since rural and remote communities often have the highest prevalence of the diseases, it is important to validate low-cost and scalable methods that can accurately screen for and diagnose SCD, SCT and β-thalassemia in these communities.

In countries like Nepal with high prevalence of β-thalassemia, it is also important to screen for β-thalassemia in addition to sickle hemoglobin, because compound heterozygous cases can be missed with common screening techniques for HbS (e.g. traditional sickling or solubility tests). Additionally, parents with heterozygous trait conditions (e.g. HbAS and HbA/β-thalassemia) with no clinical symptoms, can potentially pass the variant genes to form compound heterozygous cases in children (HbS/β-thalassemia). A study in Nepal is required to test low-cost techniques (e.g. Gazelle Hb variant test, HemoTypeSC, automated sickling test, solubility test) and compare them with the gold standard HPLC tests. The current study focuses on SCD, SCT and β-thalassemia, due to the high prevalence of HbS and β-thalassemia in Nepal. Another study (related to a different ethics application, H21-01929) will be conducted in Vancouver to validate low-cost tests for detecting SCD/SCT, and will not include participants with β-thalassemia (due to the low prevalence of β-thalassemia in Vancouver).

2.0 Study Purpose

The purpose of this study is to determine the accuracy of low-cost point-of-care techniques for screening and detecting sickle cell disease, sickle cell trait, and β-thalassemia, which will subsequently inform on feasible solutions for detecting the disease in rural, remote, or low-resource settings. One of the goals of the study is to evaluate the feasibility of techniques, such as the sickling test with low-cost microscopy and machine learning, HbS solubility test, commercial lateral-flow assays (HemoTypeSC and Sickle SCAN), and the Gazelle Hb variant test, to supplement or replace gold standard tests (HPLC or electrophoresis), which are expensive, require highly trained personnel, and are not easily accessible in remote/rural settings.

3.0 Hypothesis

We hypothesize that

1. Gazelle diagnostic device can detect β-thalassemia and SCD/SCT with an overall accuracy greater than 90%, compared with HPLC as the reference test

2. an automated sickling test (standard sickling test enhanced using low-cost microscopy and machine learning) has a higher overall accuracy than conventional screening techniques (solubility and sickling tests) to detect hemoglobin S in blood samples
3. the automated sickling test can additionally classify SCD, SCT and healthy individuals with a sensitivity greater than 90%, based on morphology changes of red blood cells, unlike conventional sickling or solubility tests that do not distinguish between SCD and SCT cases

Overall, we hypothesize that an assessment of the performance and accuracies of low-cost point-of-care techniques (automated sickling test, solubility test, lateral-flow assays, Gazelle Hb variant test) against HPLC tests will provide researchers and health workers with feasible alternative options for screening and detecting SCD, SCT and β-thalassemia in a variety of situations based on the needs of the communities and the resources available.

4.0 Objectives

Objectives specific to the current study are to:

1. Determine accuracy (sensitivity and specificity) of automated sickling test to detect HbS, compared to gold standard HPLC, and to conventional solubility test
2. Determine whether SCD, SCT and healthy individuals can be classified using the automated sickling test that leverages machine learning on images of blood films under hypoxia
3. Validate accuracy (>95% sensitivity and specificity) of lateral-flow assays (HemoTypeSC and Sickle SCAN) to detect SCD/SCT, and of Gazelle variant test to detect SCD, SCT, and β-thalassemia; and determine if low-cost techniques can potentially replace HPLC/electrophoresis tests in rural and remote settings

Long-term objectives of the overall project are to:

1. Implement trained machine learning algorithm to classify SCD, SCT and healthy individuals during screening tests in Nepal
2. Implement relevant low-cost point-of-care techniques in rural and remote communities of Nepal using insights and conclusions from current study

5.0 Endpoints/Outcomes

The study has non-clinical endpoints (endpoints not relating to how a person feels, functions or survives, but which are objective indicators of biological processes) [44], with minimal intrinsic value to the participants of the current study, since most of the participants in the study will already know the diagnosis of SCD or SCT. However, the non-clinical endpoints are strongly associated with a meaningful clinical outcome of determining optimal screening alternatives for rural and remote communities.

Primary endpoint/outcome: Characterization of accuracy (sensitivity and specificity) of low-cost techniques, such as automated sickling test (standard sickling test enhanced using low-cost microscopy and machine learning), solubility test, HemoTypeSC, Sickle SCAN, and Gazelle Hb Variant test, for detecting sickle cell disease, sickle cell trait, and β-thalassemia, based on comparison with the gold standard test (HPLC).

Secondary endpoint/outcome: Training and testing of machine learning algorithm using de-identified images collected of blood films during sickling test. Assessment of whether a microscopy / machine learning based classification approach is suitable for deployment in rural areas in Nepal.

6.0 Research Design

6.1 Research Plan and Design

The overall project has two phases:

1. **Vancouver trials (September 2022 - April 2023)**: With the help from hematologists and hematopathologists at BC Children's Hospital (BCCH) and St. Paul's Hospital (SPH), we plan to test the following screening techniques on patients with sickle cell disease (SCD) and sickle cell trait (SCT), and healthy participants:
    a. Low-cost screening
        i. Sickling test with low-cost automated microscope (such as Octopi microscope) and automated screening with machine learning
        ii. HbS solubility test
        iii. Commercial point-of-care assays (HemoTypeSC and Sickle SCAN)
        iv. Gazelle Hb variant test
    b. Gold standard test: HPLC, for determining the accuracies of low-cost screening techniques
2. **Nepal trials (October 2022 - April 2023):** With help from Dr. Rajan, CP Nepal, UBC sickle cell team, and local lab technicians and health workers, we plan to screen the communities (e.g. in Nepalgunj) using the following:
    a. Low-cost screening
        i. Sickling test with low-cost automated microscope (such as Octopi microscope) and automated screening with machine learning
        ii. Sickling test with traditional microscope (conventional manual screening used in Nepal)
        iii. HbS solubility test
        iv. Commercial point-of-care assays (HemoTypeSC and Sickle SCAN)
        v. Gazelle Hb variant test
    b. Gold standard test: HPLC, for determining the accuracies of low-cost screening techniques

This research protocol relates to the Nepal trials. There will be two ethics application submitted concurrently for the Nepal trials – one to the UBC Research Ethics Board and another to the Nepal Health Research Council.

To improve the transparency and completeness of reporting the findings (e.g. for publications resulting from this study), the checklist from the Standards for Reporting of Diagnostic Accuracy Studies or STARD 2015 will be followed [45]. The STARD 2015 guidelines was developed by experts to help reduce bias in diagnostic accuracy studies and to ensure that their reporting includes necessary information.

6.2 Duration of Study

The study is expected to commence in October/November of 2022 as soon as all regulatory requirements have been fulfilled including clinical Research Ethics Board approval. The recruitment, consent and sample analysis is expected to take 1-3 months from study commencement.

6.3 Statistical Analysis Plan

Considering HPLC as the reference test, the sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV) of the index tests or low-cost techniques (automated

sickling test, HbS solubility test, HemoTypeSC, Sickle SCAN, and Gazelle Hb variant test) will be evaluated.

For each case, the following definitions apply:

- True positive (TP): indicates a positive result for both the index test and the reference test
- True negative (TN): indicates a negative result for both the index test and the reference test
- False positive (FP): indicates a positive result for the index test, but a negative result for the reference test
- False negative (FN): indicates a negative result for the index test, but a positive result for the reference test

The following metrics will be calculated using the total number of TP, TN, FP, and FN:

$$Sensitivity = \frac{TP}{TP + FN} \times 100\%$$

$$Specificity = \frac{TN}{FP + TN} \times 100\%$$

$$PPV = \frac{TP}{TP + FP} \times 100\%$$

$$NPV = \frac{TN}{TN + FN} \times 100\%$$

The metrics (sensitivity, specificity, PPV, and NPV) for the index tests will be calculated for the following two categories:

1. **Detection analysis by hemoglobin variant** (HbA, HbS, HbC, β-thalassemia): Heterozygous or homozygous forms containing a particular hemoglobin variant in both the index test and the reference test are true positive results for that variant. Test results lacking a particular hemoglobin variant in both index and reference tests, irrespective of other variants present, are true negative results for that variant. The following hemoglobin variants will be analyzed for the index tests.
    - HbS Solubility: HbA, HbS
    - Automated sickling: HbA, HbS
    - HemoTypeSC: HbA, HbS, HbC
    - Sickle SCAN: HbA, HbS, HbC
    - Gazelle Hb Variant: HbA, HbS, HbC, β-thalassemia
2. **Detection analysis by phenotype**: In this analysis, the metrics for homozygous and heterozygous states will be calculated differently. The following phenotypes will be analyzed for the index tests.
    - Automated sickling: HbAA, HbAS, HbSS, Overall
    - HemoTypeSC: HbAA, HbAS, HbSS, (HbAC, HbSC, HbCC), Overall
    - Sickle SCAN: HbAA, HbAS, HbSS, (HbAC, HbSC, HbCC), Overall
    - Gazelle Hb Variant: HbAA, HbAS, HbSS, HbS/β-thalassemia, HbA/β-thalassemia, Hbβ/β-thalassemia, (HbAC, HbSC, HbCC), Overall
    - HbS Solubility: Not applicable, since solubility test can only identify the hemoglobin variant but not the homozygosity or heterozygosity

Any data that is excluded from the analysis, due to indeterminate results, missing data, withdrawal of consent, or any other reason, will be reported.

A total of around 120 participants – around 20 individuals with SCD (HbSS), 20 individuals with SCT (HbAS), 20 individuals with sickle cell / β-thalassemia compound heterozygous form (HbS/β-thalassemia), 20 individuals with β-thalassemia (Hbβ/β-thalassemia), 20 individuals with β-thalassemia trait or carrier form (HbA/β-thalassemia), and 20 healthy individual participants (HbAA) – will be recruited for the study.

7.0 Detailed Research Procedures

7.1 Recruitment

The following procedure will be used for recruitment:

1) Potential participants may be patients diagnosed in Bheri Hospital or Mount Sagarmatha Polyclinic, or healthy volunteers.
2) Participants will have not given prior written permission to be contacted for research purposes. The participants will first be approached for the purpose of study recruitment by someone who is within their circle of clinical care. The health care professional will approach patients directly to advise them that a study is being conducted and to enquire if the patient would like to find out more information concerning it. With permission, patients may be referred to study staff.
3) The study coordinator or delegate will discuss in person (or virtually or over the telephone) potential study participation with candidates identified. The study Informed Consent Form (ICF) will be delivered to the interested patient for review during an in-person visit.
4) If after reviewing the ICF and/or discussing it with his/her family doctor and/or friends, the patient is interested in participating in the study, the study coordinator will arrange for an appointment to further discuss the study in detail. At that visit, the SC or delegate will meet with the patient to discuss the study and review all sections of the Informed Consent Form (ICF). An investigator will be available, if needed, to discuss any questions/ concerns the patient may have of a medical nature that the SC is not qualified to answer.
5) An investigator (e.g. Dr. Rajan Pande) will be available to discuss the medical aspects of the study, and to answer any questions related to the study. However, the investigator will not obtain consent or receive the participant's answer regarding their final decision, in order to mitigate the potential for undue influence over a prospective participant. The final consent will be obtained by the study coordinator, delegate or hospital staff.
6) If the patient remains interested in participating in the study, he or she will then sign the Informed Consent Form and the screening process will commence thereafter.

7.1.1 Participants with HbS and/or β-thalassemia

All the participants with HbS and/or β-thalassemia (all participants excluding healthy volunteers) will be identified using the records of Bheri Hospital or Mount Sagarmatha Polyclinic in Nepalgunj. Records for the last 2-3 years will be accessed. Patients (and if applicable, their guardians) will be informed about the research trial and if willing to participate, informed consent will be obtained. The initial contact by delegate will happen either during in-person clinic visits or through telephone. Upon consent/assent, blood samples (3-4 mL) will be drawn for the research study (in addition to the routine bloodwork, if applicable).

7.1.2 Healthy volunteers

For recruitment of normal participants or healthy volunteers, posters/flyers will be placed on bulletin boards at the study site (Mount Sagarmatha Polyclinic).

7.2 Inclusion and Exclusion Criteria

7.2.1 Inclusion Criteria

Since the techniques evaluated in the study aims at detecting sickle cell disease (SCD), sickle cell trait (SCT), and β-thalassemia, the following number of participants will be included in the study:

- 20 individuals with SCD (HbSS)
- 20 individuals with SCT (HbAS)
- 20 individuals with sickle cell/β-thalassemia compound heterozygous form (HbS/β-thalassemia)
- 20 individuals with β-thalassemia (Hbβ/β-thalassemia)
- 20 individuals with β-thalassemia trait or carrier form (HbA/β-thalassemia)
- 20 healthy individual participants or normal participants (HbAA, participants without any known hemoglobin disorders, such as SCD, SCT or β-thalassemia)

Participants older than 1 year of age at the time of drawing blood will be eligible. Signed and dated consent or assent forms will be required by the participants or their parents/guardians.

7.2.2 Exclusion Criteria

The exclusion criteria for the study:

- Transfusion within the last 3 months
- Pregnancy

Participants who wish to withdraw from the study will also be excluded.

7.3 Consent and Assent

7.3.1 Consent process

Informed consent will be obtained by study coordinator or delegate during in-person visits. Written consent and assent forms will be provided in Nepali and will be signed and dated by participants or by guardians/parents of participants. For all participants, signed consent forms (hard copy) will be used as documentation for consent. All the consent forms will be signed in-person by pen on physical copies of the form. All participants will receive physical copies of the consent forms, while the researchers will retain physical copies of the signature pages with the corresponding version names and dates (duplicate or counterpart copies). The physical copies (duplicates) of the consent forms will be stored in a secure location in Mount Sagarmatha Polyclinic and Diagnostic Center (e.g. in Dr. Rajan Pande's office), and Dr. Rajan Pande (the investigator in Nepal) will have access to the physical copies. The consent forms with identifiable information of participants will not be transferred to UBC or Canada, and only the de-identified information (including unique participant codes) will be transferred to Canada, along with the de-identified data (e.g. images of blood films, test results, demographic information).

Assent forms and adolescent assent forms will be provided to children between 7-13 years and 14-18 years, respectively. The consent and assent forms will be stored separately from de-identified data, and the de-identified data will only be referred using unique participant codes. It will be emphasized that participation is completely voluntary and provision of care will not be affected should any patients associated with the hospital/clinic decline to participate.

7.3.2 Free, informed and ongoing consent

Participants will be clearly informed about the risks and benefits of the research, and informed about context of the current study in terms of the overall project. They will be informed that de-identified images of blood films will be used to train machine learning algorithms and publicly stored in an online database created in the Federated Research Data Repository (FRDR), an initiative by the Digital Research Alliance of Canada.. In accordance with the Declaration of Helsinki, Good Clinical Practice (GCP) guidelines and the Tri-Council Policy, a subject has the right to withdraw from the study at any point and for any reason, in which case, results and collected data will be removed and excluded from analyses.

Personal identifiers, such as names, date of birth, personal health number, etc. will not be disclosed in the study. Demographic information, such as age, sex and race, will be recorded and reported to account for and represent different distributions of hemoglobinopathies in different populations. If willing, participants can disclose their age, sex and race in a separate questionnaire. Although demographic data is available through the records of the hospital/clinic, the data for this study (age, sex and race) will only be collected through a questionnaire. The questionnaire does not collect any personal identifiers and is linked to the study ID or the unique participant codes. The data will be entered in the data collection form (Excel) and saved in password protected and encrypted computers and hard drives. To indicate the optional nature of the fields in the questionnaire, a checkbox with "Prefer not to answer" will be provided for age, sex and race. Since most of the population in the study will be from South Asian descent, this option will be provided along with “Other”, “Do not know”, and “Prefer not to answer”. The reason for collection and dissemination of demographic information will be explained in the consent forms and questionnaires.

7.4 Blood drawing and de-identification of blood samples

Certified phlebotomists will use the standard technique currently being used to draw blood at the hospital/clinic, with no special changes related to this study. A sample technique (which could vary slightly) is provided - The antecubital fossa will be identified and sterilized with an alcohol based sanitizer. After application of the tourniquet, blood will be drawn primarily from the median cubital vein. If this vein is difficult to identify or collect a sample from, other veins in the antecubital fossa will be used (cephalic vein or basilic vein). The vein will be punctured using a winged or butterfly hypodermic needle, and around 3-4 mL of blood will be collected in an evacuated container (e.g. Vacutainer) pre-loaded with anticoagulant (such as ethylenediaminetetraacetic acid or EDTA). For young children, if the veins of the antecubital fossa are not identifiable, dorsal metacarpal veins will be used. The status of the patient will be checked to confirm that bleeding has stopped. The container with the collected blood will be labelled and transported securely.

The blood samples collected will be associated with unique participant codes with no identifiable patient information. The participant code will contain details of the site of collection, date collected, and a serial number or time collected. For instance, the code ‘MSP-20220502-001’ indicates the first blood sample collected at Mount Sagarmatha Polyclinic on 02 May 2022. The study team members will be provided with only the codes, and data will be analyzed using the code as the main label or reference. Blood samples labels, images, test results, etc. will all be associated only with the unique codes, and not contain identifiable information.

Only the doctors/staff at the respective collection sites will have access to the reference list of personal identifiers (such as name, date of birth, health number) linking to the corresponding codes or de-identified data. The list will be stored in encrypted files in password protected

computers. This list will not be shared with the other study team members and the list is not required for analysis. Demographic data such as age, sex and race will be shared and reported only to properly identify the population represented in the study, and this will be clearly indicated in the consent/assent forms.

Images taken of blood will contain no identifiable information and will only reference the codes or de-identified data. Additionally, the information contained in the microscope images, which mainly includes the morphology of blood cells before or after adding reagents, cannot be linked back to any of the participants.

7.5 Sample preparation and tests

Sample preparation and tests will be performed in the hematology lab at Mount Sagarmatha Polyclinic by study team members and/or qualified hospital staff/technicians. Study team members preparing reagents and blood samples will have appropriate personal protective equipment (PPE) as determined by lab protocols. All study team members working with blood samples will have completed the necessary training required to work in the hematology lab (*e.g.* UBC Biological Safety, UBC Chemical Safety, PHSA safety courses, *etc.*). All work surfaces will be decontaminated after sample preparation, and contaminated laboratory material (e.g. pipette tips) will be disposed in a sharps/biohazard container.

7.5.1 Sickling Test

The methods used for the sickling test are as follows:

Reagent (2% sodium metabisulphite) preparation

- Add 0.2 g of sodium metabisulphite to 10 mL of distilled water in a vial/container
- Mix or stir till dissolved
- Prepare fresh every time (and use within 3 hours)

Microscope slide preparation

- Mix 1 drop of blood with 1 drop of 2% sodium metabisulphite on a microscope slide
- Add cover slip
- Incubate at room temperature for 30 minutes

Manual observation under microscope

- Examine slide under microscope (20x objective, 10x eyepiece)
- Identify as positive sickling test (SCD or SCT) if greater than 25% of the red blood cells appear sickled or deformed (Figure 1 shows sickling in an SCT individual)

Automated scanning using microscope

- Place microscope slide in slide holder of microscope unit (Octopi microscope)
- Identify the edges of the field of view for scanning
- Run program to automatically scan and capture images in the computer (translation of the stage and slide enabled by motors)
- Save images for machine learning algorithm (training/testing data set)

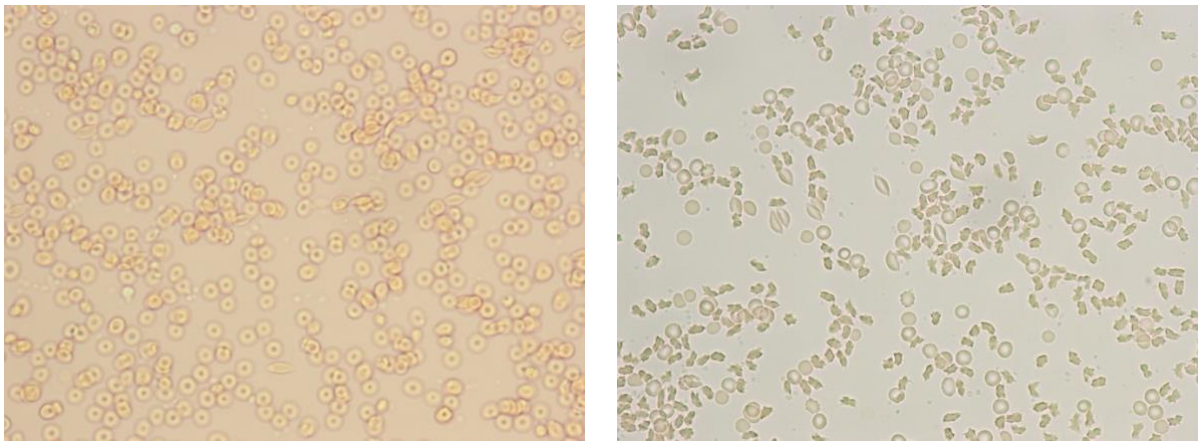

Figure 1 | Blood film containing SCT blood mixed with 2% sodium metabisulphite, showing minimal sickle cells around 10 minutes after sample preparation (left) and increased number of sickle cells around 1.5 hours after sample preparation (right); 40x objective, 10x eyepiece

7.5.2 HbS Solubility Test

The methods used for the HbS Solubility test are as follows:

Reagent preparation

- Add reagents in the following proportion (and in the order specified) [2]:
  - 1 liter of water
  - 215 g of anhydrous dipotassium hydrogen phosphate K_2HPO_4
  - 169 g of anhydrous potassium dihydrogen phosphate KH_2PO_4
  - 5 g of sodium dithionite $Na_2S_2O_4$
  - 1 g of saponin
- The prepared solution is stable for 7 days in a refrigerator

Testing procedure [2]

- Add 2 mL of reagent to three test tubes using a pipette
- Let reagent warm to room temperature
- From EDTA container, add 10 µL of packed cells to one tube
- Create positive control using 10 µL of packed cells from known SCT patient (HbAS) to second tube
- Create negative control using 10 µL of packed cells from a known normal subject (HbAA) to third tube
- Mix the blood-reagent mixture well and let stand for 5 minutes
- Ensure that the blood-reagent mixture is light pink or red (a light orange color appears if the reagent has deteriorated)
- Place tubes 2.5 cm in front of a white sheet with narrow black lines
- Compare the turbidity with positive and negative controls
- The test is positive if the solution is turbid (black lines are not observed through the solution), and is similar to the positive control, as shown in Figure 2
- The test is negative if the solution is clear (black lines are observed through the solution), and is similar to the negative control, as shown in Figure 2

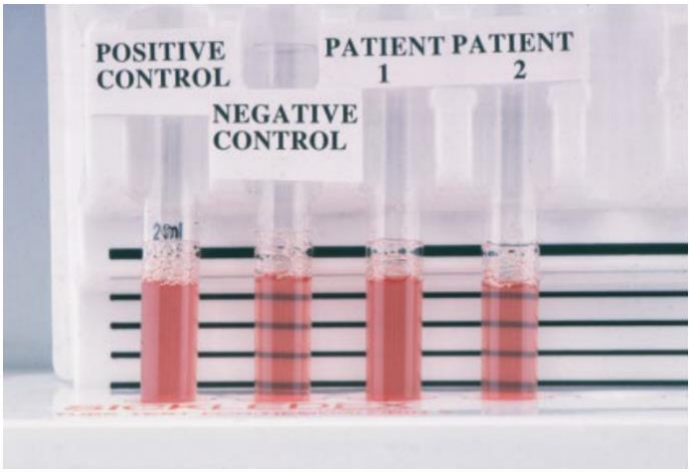

Figure 2 | Results of the HbS solubility test showing (from left to right) a positive control (turbid), a negative control (clear), a positive test (turbid) and a negative test (clear) [13]

7.5.2 HemoTypeSC

Testing procedure [46]

- Add six drops of water to test vial, using dropper pipette (Step 1 in Figure 3)
- Obtain around 1 microliter of blood sample, using fingerprick, heel-stick or EDTA tube
- Place white pad of the blood sampling device on blood sample until it absorbs a droplet of blood (Step 2 in Figure 3)
- Place blood sampling device into test vial, and swirl to mix till blood has transferred to the test vial (Step 3 in Figure 3)
- Ensure that the water has turned pink or light-red in color
- Leave the blood sampling device in the test vial
- Insert the HemoTypeSC test strip into the test vial, with the arrows pointing down (Step 4 in Figure 3)
- Wait for ten minutes (Step 5 in Figure 3)
- Remove the HemoTypeSC test strip and read the results, as shown in Figure 4

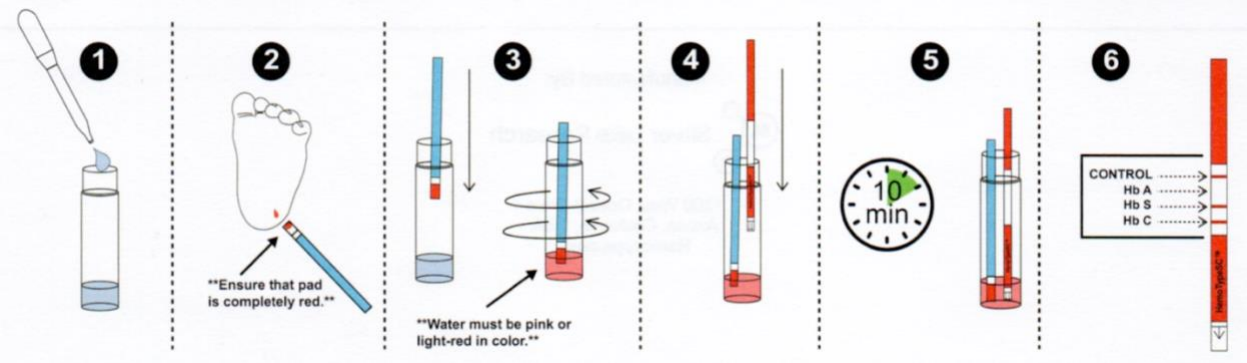

Figure 3 | Testing procedure of HemoTypeSC

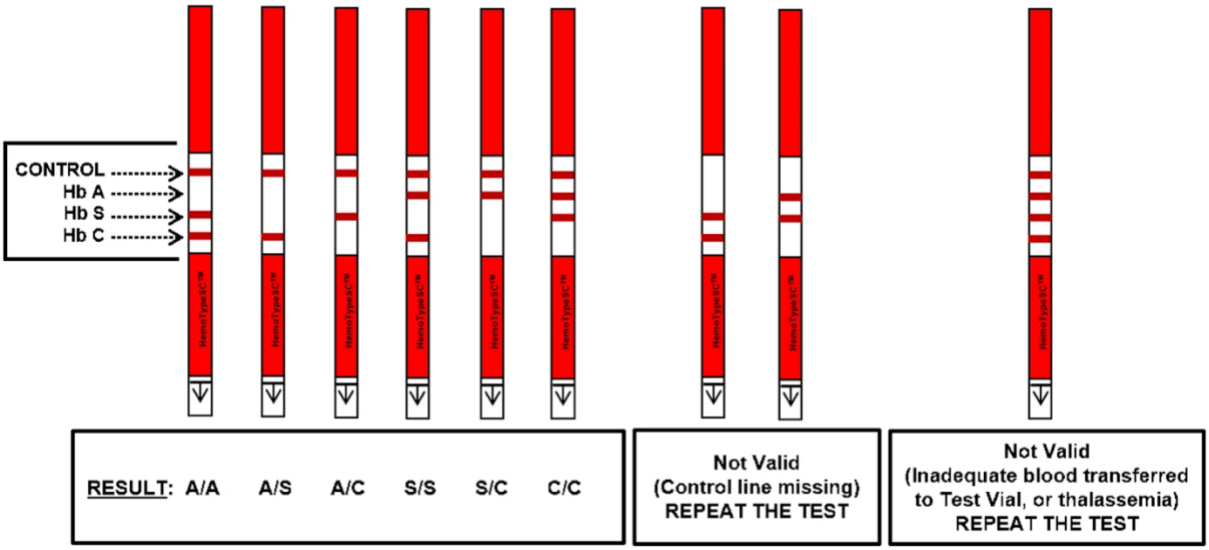

Figure 4 | Results of HemoTypeSC test strip indicating all possible combinations [46]

7.5.3 Sickle SCAN

Testing procedure

- Collect five microliters of blood, taken by fingerprick, heel-stick, or venipuncture, using the capillary sampler provided (Step 1 of Figure 5)
- Mix blood with buffer solution provided in the PreTreatment Module, to release hemoglobin by lysing red blood cells (Step 2 of Figure 5)
- Add five drops of treated sample from the PreTreatment Module into the sample inlet of the Sickle SCAN cartridge
- Wait for 5 minutes
- Read the results as shown in Figure 6

Figure 5 | Testing procedure of Sickle SCAN [47]

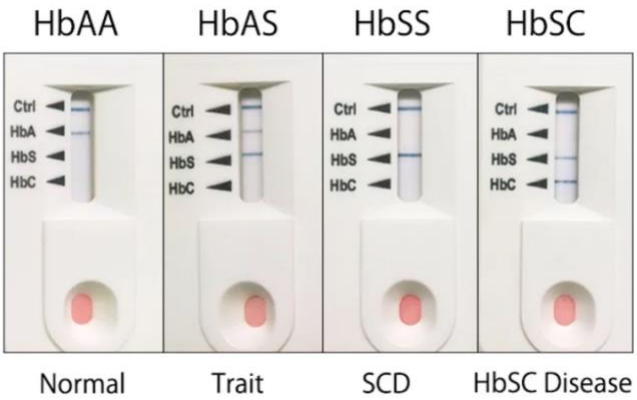

Figure 6 | Test results of Sickle SCAN showing Normal, SCT, SCD and HbSC disease cases [47]

7.5.4 Gazelle Variant Test

Testing procedure [48]

- Mix blood with marker fluid (Step 1 in Figure 7)
  - Pipette 40 µL of Gazelle Marker Fluid and 20 µL of blood into an Eppendorf tube and close the lid on the tube.
  - Vortex the solution for 20 seconds.
- Wet cartridge paper with buffer (Step 2 in Figure 7)
  - Open an Hb Variant Cartridge from the Hb Variant Test Multipack.
  - Hold the cartridge at a 35-45° angle from horizontal. Hold the pipettor at 90° to the cartridge and in one smooth motion, slowly pipette 50 µL of Gazelle Buffer into the round hole next to the cartridge slot.
  - Insert the cartridge into the stamper stand with the cartridge's notched corner facing up.
  - Wait for 60 seconds
- Apply blood sample to the Cartridge (Step 3 in Figure 7)
  - Lay the cartridge flat with the slot side up. Place the stamper stand directly over the cartridge.
  - Pipette 20 µL of the blood and marker fluid mixture onto a glass slide.
  - Gently touch the corner of a new stamper to the mixture until the stamper tip is filled from end to end.
  - Gently place the stamper with blood and marker fluid mixture into the slot on the stamper stand.
  - Place a finger on each side of the stamper stand and hold it down for five seconds.

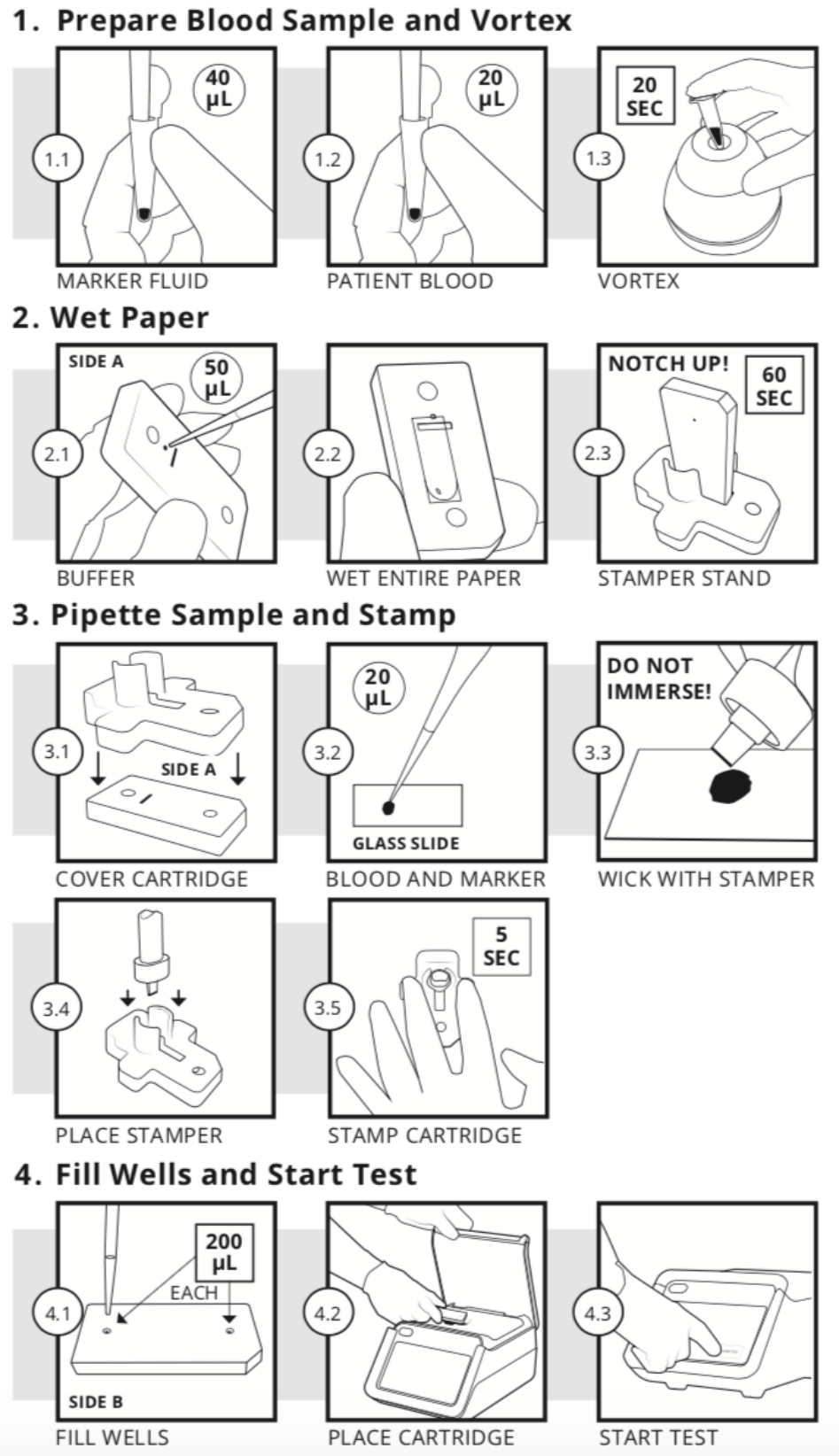

Figure 7 | Testing procedure of Gazelle Hb Variant [48]

- Fill wells and start test (Step 4 in Figure 7)
    - Remove the stamper and stamper stand from the cartridge.
    - Flip the cartridge over to SIDE B with the notched corner facing the user, and pipette 200 µL Gazelle Buffer into the left well first
    - Pipette 200 µL of buffer into the right well. For best results, angle the pipette toward the center of the cartridge.
    - Place the cartridge on the reader
- Close the lid and press START TEST on the touchscreen

7.5.5 HPLC

HPLC will be performed using the same technique used for routine tests for patients at the hospital/clinic. Each test uses approximately 5 µL of whole blood.

7.6 Equipment

7.6.1 HPLC

The HPLC instrument at Mount Sagarmatha Polyclinic is the D10 System by Bio-Rad Laboratories. Quality control is performed prior to each run or with each new cartridge as per manufacturer's recommendations.

7.6.2 Microscope

A low-cost automated microscope platform, Octopi (Open configurable high-throughput imaging platform for infectious disease diagnosis in the field), will also be used to capture images of blood films. A 20x magnification objective lens (infinity corrected) from Olympus will be used, along with a tube lens and a monochromatic camera. The microscope assembly will be equipped with motors for translation and autofocus. More details about Octopi (and related Squid) platforms can be found in https://doi.org/10.1101/684423 and https://doi.org/10.1101/2020.12.28.424613 .

7.7 Laboratory Specimen Preparation and Analysis Training

All study team members working with blood samples at the hospital/clinic will have completed the necessary safety training required to work in the hematology lab. Safety courses for members from UBC include those from UBC (Biological Safety, Chemical Safety) and Provincial Health Services Authority (Infection Prevention and Control Practices, WHMIS 2015, Occupational Health and Safety, COVID-19 Guidelines and Procedures, Respectful Workplace, Provincial Violence Prevention, *etc.*). The methodologies used for laboratory specimen preparation and analyses for the low-cost tests have been assessed by Dr. Nicholas Au (hematopathologist) and Dr. Mykola Maydan (Technical Coordinator for Special Hematology laboratory) at BC Children's hospital, and are the same as those performed for the Vancouver trials. HPLC tests will be performed by certified staff at the hospital/clinic. The results for laboratory tests (*e.g.* HPLC) will be interpreted by doctors at the corresponding site.

7.8 Neural Network Training and Testing

Data manipulation and analysis will be done through Python. All software to be utilized is open-source and does not have commercial implications. In order to optimize the screening algorithm, multiple neural network configurations will be tested and compared. All the models tested for classification will use existing pre-trained models that have applications in image classification.

In each case, the acquired images will be split into three groups, consisting of training data (around 80%), validation data, (around 10%) and testing data (around 10%). The data will be used to train the neural network using a pre-trained supervised machine learning model, whereby the images are input and the parameters of the network are optimized to classify images by hemoglobin phenotype. The test data will test the performance of the trained network on images that have not been used for training. We will use the neural networks most commonly used for medical image classification, such as EfficientNet, ResNeXt, ResNet-50, DenseNet, and Xception. When possible, training will occur locally on the laboratory computer at UBC. For networks that require additional computing power, UBC ARC Sockeye or Compute Canada will be used.

All the machine learning algorithms will only include de-identified data, such as images of blood films and hemoglobin phenotype (such as sickle cell trait or HbAS). Personal identifying information, such as participant names, will not be associated with any of the data used for analysis or stored online.

Cloud computing facilities, such as those from Compute Canada, will be utilized to train and test machine learning algorithms on the final dataset of images. Compute Canada is a cloud computing service provided by Digital Research Alliance of Canada (the Alliance) to Canadian research institutions. The servers and clusters (e.g. Cedar cluster located in BC) will be accessed using Secure Shell Protocol (SSH). For the analysis (before sharing the data publicly), de-identified images may be stored securely in UBC servers or in servers for Compute Canada.

The de-identified data (such as images of blood films and associated de-identified data like hemoglobin phenotype) will be deposited in an online public repository, such as Federated Research Data Repository.

For this study, the machine learning algorithm will not be used to diagnose new patients, and the ground truth will be derived from results from the gold standard test (HPLC). The images of blood cells will be trained/tested using results from HPLC as the ground truth (e.g. diagnosis of sickle cell trait).

7.9 De-identified images shared in online repository

De-identified data (images of blood films from a microscope and associated documentation) will also be deposited in an online public repository, such as the Federated Research Data Repository (FRDR). FRDR is a service of the Digital Research Alliance of Canada (Alliance), a not-for-profit organization that supports digital research infrastructure in Canada. FRDR is hosted on national infrastructure, managed and administered by the Digital Research Alliance of Canada. Production Virtual Machines (VMs) are hosted on clusters at the University of Waterloo (Graham) and Simon Fraser University (Cedar). Data are mirrored across sites, with only one site active at a time, and failover to the other site requiring a manual process. Archival copies of the files are stored on a cluster at the University of Toronto (Niagara), maintained in agreement with the University of Toronto Libraries and SciNet (the supercomputer centre at the University of Toronto). All servers are located in Canada.

FRDR is a curated, general-purpose data repository available to Canadian researchers. All deposits undergo review by a member of the FRDR curation team to ensure compliance with the repository terms of use. An unsigned copy of the consent form that was provided to study participants will be reviewed to confirm permission was granted to share files in a public repository such as FRDR, and files will be scanned for any potentially identifying information.

When datasets are published in FRDR, a SHA256 hash is created for each file. This will be used to verify file integrity over time, and used to verify the successful transfer and ingest of files for archival processing. All images are digital, and will be available through FRDR's online portal at https://www.frdr-dfdr.ca/repo/. The images will be in the FRDR repository for a minimum of 10 years, and will undergo long term preservation using archival processing.

FRDR policies are available online, including a data submission policy [https://www.frdr-dfdr.ca/policies/en/data_submission/], an access and reuse policy [https://www.frdr-dfdr.ca/policies/en/access_reuse/], and a data retention and deaccession policy [https://www.frdr-dfdr.ca/policies/en/data_retention/].

8.0 Study Schedule

Screening: Potential participants will be approached by investigators/study delegates who will explain the rationale, goals and risks and benefits of study participation. If subjects acknowledge and confirm voluntary participation in the study, then the informed consent form will be signed and the participant's enrolment will commence at the time of next blood collection.

Blood Draws: Participants who have been consented within the past 30 days will arrive at either Mount Sagarmatha Polyclinic for drawing blood as part of their routine visit (e.g. for SCD or SCT participants) or a visit for the study (e.g. for normal participants).

Tests: The HPLC tests will be performed at Mount Sagarmatha Polyclinic. The low-cost tests will be performed at the hematology lab at the hospital/clinic. Images of blood films for the sickling tests will be taken at the hematology lab using the Octopi microscope assembly.

Analyses: The machine learning algorithms will be trained and tested after collecting the images for blood films of the sickling test. The analyses will be performed at UBC. The performance of the algorithms will be tested with the data collected during the study, and at the end of the study (after completion of all tests). The statistical analyses will be performed after all the data has been collected.

Reporting results: After completing the study, the results will be published either as a separate report or together with the results from the Vancouver study.

9.0 Reporting of Adverse Events (AE's), Severe Adverse Events (SAE's) and Unanticipated Problems (UP's)

It is unlikely that this study will result in significant harm towards participants. However, phlebotomy is associated with the following risks, most of which can be managed:

1. Excessive bleeding
2. Fainting or lightheadedness
3. Hematoma or blood accumulating under the skin
4. Infection from blood borne pathogens
5. Multiple punctures to locate veins

The probability of these risks to participants is usually low, if best practices of phlebotomy are followed. All individuals performing phlebotomy will have received the appropriate training, certificates of approval, and follow best practice procedures to minimize these risks. Appropriate care will be provided on an individual basis, if participants experience any potential harm identified above.

Any unanticipated problems will be recorded and reported to the REB. Serious adverse events will be reported using the request for acknowledgement form as soon as reasonably possible, but in any case, within seven days from their occurrence.

10.0 Documentation and Record Keeping

The investigators and delegates will be responsible for maintaining detailed records of all study participants that undergo screening. This responsibility extends to medical records and laboratory testing. De-identified images and data will be stored in external hard drives and local drives of laboratory computers used by study members to perform analysis. The external hard drive will be stored at the study team member's office during the study, and at the Principal Investigator's office after the analysis for the study has completed. The external hard drives and computers will be password protected. External hard drives storing any data will be encrypted. De-identified images will be analyzed in Canada, and appropriate data transfer agreement will be obtained if necessary.

11. Ethical Considerations

UBC Research Ethics Board and Nepal Health Research Council (NHRC) Research Ethics Board approval of the protocol and all study documents will be sought prior to any patient recruitment. The required certificates of institutional approval will be obtained prior to commencing recruitment of participants. The study will comply with the principles outlined in ICH Good Clinical Practices. REB approval will also be sought for any required protocol amendments.

12. Feasibility

Approximately 800 patients with SCA have been identified at Bheri Hospital and Mount Sagarmatha Polyclinic, and more with participants with sickle cell trait or β-thalassemia forms have been identified. Since the Bheri Hospital and Mount Sagarmatha Polyclinic are have HPLC machines, most patients from Nepalgunj and neighboring areas come to the hospital for diagnosis. Since Bheri hospital provides the diagnosis for free for SCD patients, the HPLC machine is relatively busy. The HPLC machine at Mount Sagarmatha Polyclinic is not as busy as that in Bheri hospital, which makes it more feasible to run the study at Mount Sagarmatha Polyclinic.

The hospital/clinic keeps a record of all patients, their diagnoses, and their contact information. As the number for each participant category is relatively small (i.e. 20 participants for each category), recruitment is feasible. We do not feel that additional participation in this study would be burdensome to patients and staff. We also plan to only recruit 20 healthy volunteers. We expect our recruitment efforts will assist in achieving this number.

13. Limitations

Many SCD and β-thalassemia patients will be undergoing treatment and taking medications, which could potentially affect the interpretability of results of the sickling test. Nevertheless, blood films of patients with treatment are expected to include sickled cells during the sickling test. The analyses (*e.g.* machine learning training) will include information of whether or not patients have had treatment or medication, and any differences in results due to the treatment will be reported.

The population of patients in Nepalgunj are mostly of South Asian descent with possibly different haplotype than that of patients elsewhere (*e.g.* in Africa). This may have some

implications in differentiating SCD vs. SCT elsewhere, if the machine learning algorithm is only trained using a population of different haplotypes. To improve the accuracy of the machine learning classification, it may be required to add data from the population being tested.

In addition to the homozygous HbSS form of SCA, other heterozygous forms of β-thalassemia (with HbS) could lead to similar levels of sickling. It is unclear whether these can be differentiated in the sickling test. Additionally, the lateral flow assays (HemoTypeSC and Sickle SCAN) do not detect β-thalassemia.

14. Originality, Clinical and Scientific Importance

Low-cost point-of-care techniques to detect SCD, SCT, and β-thalassemia are essential to improving patient outcomes in low-income and middle-income countries, where the burden of the disease is the highest. An evaluation of the performance and feasibility of multiple low-cost techniques, which the current study aims to do, is required to inform health workers and researchers of the most optimal solutions for detecting the disease in remote/rural communities.

15.0 References

[1] F. B. Piel, M. H. Steinberg, and D. C. Rees, “Sickle Cell Disease,” *N. Engl. J. Med.*, vol. 376, no. 16, pp. 1561–1573, Apr. 2017.

[2] B. J. Wild and B. J. Bain, “Investigation of Variant Haemoglobins and Thalassaemias,” in *Dacie and Lewis Practical Haematology*, Twelfth Ed., Elsevier, 2017, pp. 282–311.

[3] D. C. Rees, T. N. Williams, and M. T. Gladwin, “Sickle-cell disease,” *Lancet*, vol. 376, no. 9757, pp. 2018–2031, Dec. 2010.

[4] R. E. Ware, M. de Montalembert, L. Tshilolo, and M. R. Abboud, “Sickle cell disease,” *Lancet*, vol. 390, no. 10091, pp. 311–323, Jul. 2017.

[5] G. J. Kato *et al.*, “Sickle cell disease,” *Nat. Rev. Dis. Prim.*, vol. 4, no. 1, p. 18010, Jun. 2018.

[6] P. Adepoju, “Ghana takes on sickle-cell disease,” *Lancet*, vol. 395, no. 10222, p. 402, 2020.

[7] J. Gyamfi *et al.*, “Evidence-based interventions implemented in low-and middle-income countries for sickle cell disease management: A systematic review of randomized controlled trials,” *PLoS One*, vol. 16, no. 2, p. e0246700, Feb. 2021.

[8] S. Simpson, “Sickle cell disease: a new era,” *Lancet Haematol.*, vol. 6, no. 8, pp. e393–e394, Aug. 2019.

[9] I. Odame, “Perspective: We need a global solution,” *Nature*, vol. 515, no. 7526, pp. S10–S10, Nov. 2014.

[10] Nepal Health Research Council, “Annual Report,” 2018.

[11] R. M. Shrestha, R. Pandit, U. K. Yadav, R. Das, B. K. Yadav, and H. C. Upreti, “Distribution of Hemoglobinopathy in Nepalese Population,” *J. Nepal Health Res. Counc.*, vol. 18, no. 1, pp. 52–58, Apr. 2020.

[12] A. Shrestha, V. Pant, K. Gautam, D. Pyakurel, and S. Pradhan, “Detection of Hemoglobinopathies by HPLC in a Referral Clinical Laboratory in Nepal,” *Nepal. Med. J.*, vol. 3, no. 1, pp. 306–308, Jun. 2020.

[13] B. J. Bain, Ed., *Haemoglobinopathy Diagnosis*. Oxford, UK: Blackwell Publishing Ltd, 2006.

[14] A. L. Okwi, W. Byarugaba, A. Parkes, and M. Ocaido, “The Reliability of Sickling and Solubility Tests and Peripheral Blood Film Method for Sickle Cell Disease Screening at District Health Centers in Uganda,” *Clin. Mother Child Heal.*, vol. 7, pp. 1–5, 2010.

[15] V. N. Tubman and J. J. Field, “Sickle solubility test to screen for sickle cell trait: what’s the harm?,” *Hematology*, vol. 2015, no. 1, pp. 433–435, Dec. 2015.

[16] J. B. Kakou Danho, Y. N. Atiméré, D. Koné, D. D. Yéo, and L. Couitchéré, “Feasibility Study of the ‘HemoTypeSC’ Test for the Rapid Screening of Sickle Cell Disease in Côte D’Ivoire,” *Adv. Hematol.*, vol. 2021, pp. 1–7, Mar. 2021.
[17] C. Steele *et al.*, “Point-of-care screening for sickle cell disease in low-resource settings: A multi-center evaluation of HemoTypeSC, a novel rapid test,” *Am. J. Hematol.*, vol. 94, no. 1, pp. 39–45, Jan. 2019.
[18] M. B. Mukherjee *et al.*, “Multicenter Evaluation of HemoTypeSC as a Point-of-Care Sickle Cell Disease Rapid Diagnostic Test for Newborns and Adults Across India,” *Am. J. Clin. Pathol.*, vol. 153, no. 1, pp. 82–87, Jan. 2020.
[19] O. Nnodu *et al.*, “HemoTypeSC, a low-cost point-of-care testing device for sickle cell disease: Promises and challenges,” *Blood Cells, Mol. Dis.*, vol. 78, no. February, pp. 22–28, Sep. 2019.
[20] J. Kanter, M. J. Telen, C. Hoppe, C. L. Roberts, J. S. Kim, and X. Yang, “Validation of a novel point of care testing device for sickle cell disease,” *BMC Med.*, vol. 13, no. 1, p. 225, Dec. 2015.
[21] A. Y. Segbena *et al.*, “Diagnostic accuracy in field conditions of the sickle SCAN® rapid test for sickle cell disease among children and adults in two West African settings: the DREPATEST study,” *BMC Hematol.*, vol. 18, no. 1, p. 26, Dec. 2018.
[22] P. T. McGann, B. A. Schaefer, M. Paniagua, T. A. Howard, and R. E. Ware, “Characteristics of a rapid, point-of-care lateral flow immunoassay for the diagnosis of sickle cell disease,” *Am. J. Hematol.*, vol. 91, no. 2, pp. 205–210, Feb. 2016.
[23] O. E. Nnodu *et al.*, “Implementing newborn screening for sickle cell disease as part of immunisation programmes in Nigeria: a feasibility study,” *Lancet Haematol.*, vol. 7, no. 7, pp. e534–e540, 2020.
[24] K. de Haan *et al.*, “Automated screening of sickle cells using a smartphone-based microscope and deep learning,” *npj Digit. Med.*, vol. 3, no. 1, p. 76, May 2020.
[25] S. Ilyas, M. Sher, E. Du, and W. Asghar, “Smartphone-based sickle cell disease detection and monitoring for point-of-care settings,” *Biosens. Bioelectron.*, vol. 165, no. July, p. 112417, Oct. 2020.
[26] C. D’Costa *et al.*, “Differential sensitivity to hypoxia enables shape-based classification of sickle cell disease and trait blood samples,” *medRxiv*, 2020.
[27] J. S. Cybulski, J. Clements, and M. Prakash, “Foldscope: Origami-Based Paper Microscope,” *PLoS One*, vol. 9, no. 6, p. e98781, Jun. 2014.
[28] N. A. Switz, M. V. D’Ambrosio, and D. A. Fletcher, “Low-cost mobile phone microscopy with a reversed mobile phone camera lens,” *PLoS One*, vol. 9, no. 5, 2014.
[29] P. Purwar, S. Han, Y. Lee, B. Saha, T. Sandhan, and J. Lee, “High-resolution cost-effective compact portable inverted light microscope,” *J. Microsc.*, vol. 273, no. 3, pp. 199–209, Mar. 2019.
[30] H. Li *et al.*, “Squid : Simplifying Quantitative Imaging Platform Development and Deployment,” 2020.
[31] C. Stewart and J. Giannini, “Inexpensive, open source epifluorescence microscopes,” *J. Chem. Educ.*, vol. 93, no. 7, pp. 1310–1315, 2016.
[32] B. Diederich *et al.*, “A versatile and customizable low-cost 3D-printed open standard for microscopic imaging,” *Nat. Commun.*, vol. 11, no. 1, p. 5979, Dec. 2020.
[33] J. P. Sharkey, D. C. W. Foo, A. Kabla, J. J. Baumberg, and R. W. Bowman, “A one-piece 3D printed flexure translation stage for open-source microscopy,” *Rev. Sci. Instrum.*, vol. 87, no. 2, p. 025104, Feb. 2016.
[34] M. N. Hasan *et al.*, “Paper-based microchip electrophoresis for point-of-care hemoglobin testing,” *Analyst*, vol. 145, no. 7, pp. 2525–2542, 2020.
[35] S. Ilyas, A. E. Simonson, and W. Asghar, “Emerging point-of-care technologies for sickle cell disease diagnostics,” *Clin. Chim. Acta*, vol. 501, no. July 2019, pp. 85–91, Feb. 2020.

[36] P. K. Das, S. Meher, R. Panda, and A. Abraham, “A Review of Automated Methods for the Detection of Sickle Cell Disease,” *IEEE Reviews in Biomedical Engineering*, vol. 13. Institute of Electrical and Electronics Engineers, pp. 309–324, 2020.
[37] M. Xu, D. P. Papageorgiou, S. Z. Abidi, M. Dao, H. Zhao, and G. E. Karniadakis, “A deep convolutional neural network for classification of red blood cells in sickle cell anemia,” *PLoS Comput. Biol.*, vol. 13, no. 10, pp. 1–27, 2017.
[38] L. Alzubaidi, M. A. Fadhel, O. Al-Shamma, J. Zhang, and Y. Duan, “Deep Learning Models for Classification of Red Blood Cells in Microscopy Images to Aid in Sickle Cell Anemia Diagnosis,” *Electronics*, vol. 9, no. 3, p. 427, Mar. 2020.
[39] B. Javidi, A. Markman, S. Rawat, T. O’Connor, A. Anand, and B. Andemariam, “Sickle cell disease diagnosis based on spatio-temporal cell dynamics analysis using 3D printed shearing digital holographic microscopy,” *Opt. Express*, vol. 26, no. 10, p. 13614, May 2018.
[40] G. A. Barabino, M. O. Platt, and D. K. Kaul, “Sickle Cell Biomechanics,” *Annu. Rev. Biomed. Eng.*, vol. 12, no. 1, pp. 345–367, Jul. 2010.
[41] N. Z. Piety *et al.*, “A Paper-Based Test for Screening Newborns for Sickle Cell Disease,” *Sci. Rep.*, vol. 7, Apr. 2017.
[42] S. M. Knowlton *et al.*, “Sickle cell detection using a smartphone,” *Sci. Rep.*, vol. 5, no. 1, p. 15022, Dec. 2015.
[43] M. Marchand *et al.*, “The Assessment and Sustainable Management of Sickle Cell Disease in the Indigenous Tharu Population of Nepal,” *Hemoglobin*, vol. 41, no. 4–6, pp. 278–282, Nov. 2017.
[44] C. McLeod, R. Norman, E. Litton, B. R. Saville, S. Webb, and T. L. Snelling, “Choosing primary endpoints for clinical trials of health care interventions,” *Contemp. Clin. Trials Commun.*, vol. 16, no. November, p. 100486, 2019.
[45] J. F. Cohen *et al.*, “STARD 2015 guidelines for reporting diagnostic accuracy studies: Explanation and elaboration,” *BMJ Open*, vol. 6, no. 11, pp. 1–17, 2016.
[46] S. L. Research, “HemoTypeSC: Rapid Detection of Hemoglobins A, S, and C in Just a Droplet of Blood.” [Online]. Available: https://www.hemotype.com/. [Accessed: 26-Jul-2021].
[47] Biomedomics, “Sickle SCAN.” [Online]. Available: https://www.biomedomics.com/products/hematology/sicklescan/. [Accessed: 26-Jul-2021].
[48] H. Health, “Gazelle Diagnostic Device.” [Online]. Available: https://hemexhealth.com/products. [Accessed: 30-Aug-2021].