Rapid, Affordable and Efficient Screening of Multiple Blood Abnormalities Made Possible Using an Automated Tool for MALDI-ToF Spectrometry Analysis

Ricardo J. Pais [1], **Christian Jardine** [1], **Raminta Zmuidinaite** [1], **Jonathan Lacey** [1], **Steve Butler** [1] **and Ray Iles** [1,2,*]

[1]   MAP Sciences Ltd., Priory Business Park, Bedford I-lab Stannard Way, Bedford MK44 3RZ, UK; ricardojorge22@gmail.com (R.J.P.); Christian.Jardine@mapsciences.com (C.J.); Raminta.Zmuidinaite@mapsciences.com (R.Z.); Jon.Lacey@mapsciences.com (J.L.); Steve.Butler@mapsciences.com (S.B.)

[2]   Dean's Office, College of Health Sciences, Abu Dhabi University, P.O. Box 59911, Abu Dhabi, UAE

*   Correspondence: raymond.iles@adu.ac.ae

Featured Application: **Population screening for blood abnormalities suggestive of haemoglob inopathies, thalassemias and pre-diabetes/diabetes in national health care control programs of Asian, Middle East and African countries.**

Abstract: Screening programs for genetic and metabolic diseases such as haemoglobinopathies, thalassemias and diabetes are a worldwide problem that faces economic and technological limitations. This is mainly because genetic and metabolic tests are too expensive and time consuming to be implemented. MALDI-ToF mass spectrometry is a rapid and affordable high throughput technique with diagnostic potential for these diseases but still constrained by the timing and complexity of data analysis. To overcome this technological limitation, we developed a fully automated software solution in our MALDI-ToF instrument towards the detection of haemoglobinopathies, thalassemias and diabetes on one blood card sample. The software was tested for its efficiency and accuracy on 171 blood samples rendering 30-fold faster analysis with less bias and rounding errors in comparison with the manual approach. In this study, we identified the variability associated with the disease biomarkers in healthy individuals and successfully applied predictive models to detect blood abnormalities. Taken together, we demonstrated in this study that population screening of multiple blood disorders is made possible using MALDI-ToF technology in combination with automated software tools.

Keywords: MALDI-ToF mass spectrometry; software applications; clinical diagnostic; blood disorders; population screening

1. Introduction

Haemoglobinopathies and thalassemias are inherited genetic mutations in the α and β subunits of haemoglobin that have been classified worldwide as endemic diseases and frequently observed in Asia, Middle East and African populations [1]. These genetic alterations lead to abnormal expression and structural changes in haemoglobin, often resulting in anemia and has been estimated that 5% of the population suffers from its symptoms [1,2]. The diagnostic of these diseases is based on expensive and time-consuming genetic tests that are not affordable for screening large populations, making this a huge problem for national health care systems [3–5]. Type 2 diabetes is also another worldwide healthcare problem that most countries face, based on the increased number of cases over the past years suggesting effective screening programs for early detection are required [6,7]. This early stage of the

disease is called pre-diabetes and preventive medicine is considered advantageous to reduce the burden of diabetes in the healthcare systems [8]. So far, rapid, effective and affordable methodologies towards screening of haemoglobinopathies, thalassemias and pre-diabetes detection in large populations are still problematic due to economic and technical constraints [8,9].

Matrix assisted laser desorption ionization (MALDI) time of flight (ToF) mass spectrometry (MS) is a direct, rapid and affordable analytical method already used in clinical microbiology for the pathogenic bacteria identification [10–12]. This methodology has been successfully implemented in clinical laboratories with demonstrated effective diagnostic power and substantial reduction in the costs per test [13–15]. Several studies have shown that the accuracy of MALDI-ToF MS allows the detection of chemical modification of biomolecules in complex mixtures such as urine and blood [16–19]. The application of MALDI-ToF for the detection of haemoglobinopathies in blood was recently demonstrated by Iles and Nasse [18]. Their proposed method has the potential of identifying not only haemoglobinopathies but also α-thalassemia, β-thalassemia, and pre-diabetes in a single sample run using a single drop of blood. This is based on the detection and quantification of globin subunits (e.g., α and β) and their correspondent glycated fractions in the MS as biomarkers for these diseases.

MALDI-ToF MS has evolved towards an ultra-fast high-throughput technology, which has the potential to be applied to the screening of large populations [18,20,21]. However, interpretation of the quality of the MS and quantification of biomarkers for the correct diagnosis of a particular disease requires highly specialized technicians. Additionally, this can insert human errors, bias, extra costs in human resources and it is time consuming, which prevents MALDI-ToF to reach its full potential as a rapid and affordable high-throughput technology. Smart and fully automated software that optimizes time costs while also increasing accuracy can provide a solution for the implementation of MALDI in clinical laboratories. Previously, we developed an automated bioinformatic workflow that successfully pre-processes raw data, performs accurate quality control decisions and analyzes large data sets from urine MS generated by MALDI-ToF [22]. We further adapted these workflows and developed a user-friendly software tool (HbAnalizer), which was designed to perform fully automated detection of abnormalities in MS of human blood in large data sets. In this work, we present a software application as a solution for the screening of haemoglobinopathies, thalassemias and pre-diabetes in clinical laboratories.

2. Materials and Methods

2.1. Blood Samples

A single drop of fresh human blood was collected from a total of 58 assorted volunteers from the UK in November 2018. All individuals provided written informed consent to participate in the study declaring their age, ethnicity, sex and relative health status. Assorted samples are composed of 42 males and 17 females. The minimum age of volunteers was 18 and the maximum 72 years old, with an average age of 41 years old.

2.2. Sample Preparation

A pinprick of fresh blood was spotted on a DBS card (SLS) and allowed to fully dry. A 6 mm single hole office punch was used to extract the blood-stained DBS card and soaked in 3 ml mass spectrometry grade water (Romil, UK) for two hours. The 6 mm single hole office punch was cleaned with methanol (Romil, UK) between each set. A further two doubling dilutions were performed on each sample provided, constituting the final dilutions of 1:500, 1:1000 and 1:2000 as estimated by measuring the volume of haemoglobin. One microliter of each dilution of each sample was plated on a MALDI-ToF stainless steel target plate with sinapinic acid (Sigma-Aldrich, St. Louis, MO, USA) matrix. Plate samples were allowed to fully air dry.

2.3. Mass Spectrometry

Mass spectra were acquired with a MALDI 8020 (Shimadzu, UK) mass spectrometer set in a linear, positive mode. The benchtop instrument was equipped with 200 Hz solid state laser (λmax = 355 nm) and pulse extraction of 23,000 nanoseconds was performed. A total of 4000 shots in 200 positions of a single plate well were obtained and the raw spectra were computed and saved in mzML files. The mass spectrometer was externally calibrated using Apomyoglobin (16,952 Da) (ProteoMass, Sigma-Aldrich) for both singly and doubly charged ions. The laboratory SOPs in place ensure that the data acquisition is always standardized minimize any potential batch effects. The extremely high repetition number and raster patterning across samples ensures that the data recorded is an average for the sample and minimizes any plate to plate variation.

2.4. Quality Control

Mass spectra were checked for quality to produce robust results based on identifying minimal data features and possible interferents on the mass region between 7200 *m/z* and 8500 *m/z*, corresponding to the doubly charged ions. First, we checked for the presence of α-globin and β-globin as well-defined peaks within plausible *m/z* ranges that are characteristic for these proteins as a minimum requisite. At this stage, all spectra that did not have both α-globin and β-globin main peaks were rejected and no further analysis was performed. Additionally, we have checked for the characteristic secondary peaks of α-globin and β-globin forms complexed with sinapinic acid, which are expected to be located ~83 *m/z* from the correspondent free form. Next, we checked for baseline interference by calculating the absolute deviation between the minimum intensity of the two extreme regions (7200–7400 *m/z* and 8300–8500 *m/z*), which is expected to be close to the y-axis in a good quality MS (<1% of maximum intensity value). For the identification of possible contaminant interference, we checked for additional well-defined peaks close to the α-globin and β-globin detected peaks. Furthermore, data normalization is implemented within the final process, so any slight batch variation does not affect the analysis.

2.5. MS Biomarkers for Blood Abnormalities

Haemoglobin subunits α-globin, β-globin and respective glycated forms were identified on the MS of blood samples using the patented method for detecting several abnormalities, described by Iles et al. [23]. According to this methodology, we calculated the distance between the detected positions of α-globin and β-globin peaks (Dαβ) as a way to detect haemoglobinopathies [18]. For the identification of α and β thalassemias, we used the calculated ratio between β-globin and α-globin (Rβα). For pre-diabetes, we used as biomarker the estimated percentage of glycated α-globin (αGlc) in the MS of blood according to the procedure described by Iles et al. [23].

2.6. Scoring the Detection of Blood Abnormalities

Scores for the detection of blood abnormalities were computed based on the estimated probability of a given level of a biomarker in a blood sample belonging to a statistical model of the disease. All models assumed that the probability of having a particular disease follows a Gaussian distribution of biomarker values. Model parameters are described by the expected value (μ) as the threshold for the biomarker and its associated standard deviation (δ), see Table 1 for values and references. Scoring functions were derived from the probability density function by normalizing its values with the maximum probability value, considering only the relevant tails. Thus, for the case of haemoglobinopathies detection, we considered both tails of the distribution (Equation (1)), where Xi the obtained Dαβ for blood sample *i*.

$$Score = 1 - e^{\frac{(Xi-\mu)^2}{2\delta^2}} \tag{1}$$

For α thalassemia and diabetes, we modelled disease detection (Equation (2)) by considering only the left side of the Gaussian distribution and assuming maximum probability when the respective biomarker value (Xi) of the sample is higher than the threshold value for the reported for the disease biomarker.

$$Score = \begin{cases} e^{\frac{(Xi-\mu)^2}{2\delta^2}}, Xi \leq \mu \\ 1, Xi > \mu \end{cases} \tag{2}$$

For β thalassemia, we modelled the detection based on the probability of not being in normal ranges for R$\beta\alpha$, only considering the left side of the biomarker variation (Equation (3)). In this case, we assumed minimum probability when the respective biomarker value (Xi) of the sample is higher than the threshold value for the disease biomarker.

$$Score = \begin{cases} 1 - e^{\frac{(Xi-\mu)^2}{2\delta^2}}, Xi \leq \mu \\ 0, Xi > \mu \end{cases} \tag{3}$$

Table 1. Model parameters and references used for the detection of blood abnormalities. *m/z* values indicate the distance between α- and β-peaks.

Disease Detection	Biomarker	μ Value	Ref.	δ Value	Ref.
Haemoglobinopathies	D$\alpha\beta$	368.2 *m/z*	[18,23] [1]	6 *m/z*	MS data [2]
α Thalassemia	R$\beta\alpha$	90%	[23] [1]	15%	MS data [2]
β Thalassemia	R$\beta\alpha$	51%	[23] [1]	25%	MS data [2]
Diabetes	αGlc	10%	[24,25] [1]	3%	[24,25] [1]

[1] Adjusted value from reference. [2] Estimated to fit the MS of normal individuals.

2.7. Manual Data Analysis

Mass spectra from blood samples were manually interpreted for quality using the free software tool mMass version 5.5.0 by performing a visual inspection for noise and spectra quality of Hb variants peaks in a zoomed-in mass region between 7200 *m/z* to 8500 *m/z* [26]. For analysis of good quality spectra, we smoothed data with a 10 *m/z* window over 5 cycles with Savitzky-Golay method and collected the values of the peak positions and intensities using the mMass processing and analytical tools [27]. All calculations of biochemical parameters for detection of blood abnormalities was performed using Microsoft Excel.

2.8. Automated Data Analysis

For a fully automated quality assessment and biomarkers detection on MS of blood samples, we used the MAP Sciences HbAnalyser beta version 1 software tool (HbAnalyser), developed in our lab. The software tool was developed in Python version 3.7 and operates on Microsoft Windows compatible computers, with an operational requirement of 2587 Kilobytes of RAM. This software was developed using wxPython 4.0 toolkit for generating a user-friendly interface (Figure 1). We implemented two tabs on the menu bar, one for single sample analysis and another for multiple sample analysis (see Supplementary Video for a full software demonstration). The software was written to be compatible with the mzML file system of Shimadzu instruments. For a single sample analysis with this tool, the user chooses a single file location and simply runs the process button to show the processed spectra and computed biochemical parameters for detection of blood abnormalities on display (Figure 1) with optional saving results in CSV or MS image. For running multiple samples systematically, the user chooses the folder containing all files and the quality control status for each sample systematically appears on the display while running the process (Figure 1), whereas the biochemical parameters for each sample detection of blood abnormalities and MS are automatically saved in CSV and MS as image.

In this software, we implemented algorithms adapted from [22] to systematically normalize, smooth, estimate noise, and peak detection/selection. For computing all parameters for detection of blood abnormalities (see Section 2.5), we developed an algorithm according to the procedure described by Iles et al. [23]. Simultaneously, the algorithm generated two computed values for each parameter analyzed; one that is based on the intensity value retrieved from expected location (Theoretical), and another based on the peak detected nearest to the theoretical peak position (Correct position). We programmed the algorithm such as the computed value of a parameter is retrieved from the theoretical location only if the method did not find a well-defined peak in the neighborhood of the theoretical location. Automated quality control decision was implemented in the tool with an algorithm adapted from the one used in [22], which checks all typical MS requisites of blood (see Section 2.4) and prints to display the summary of quality checking with possible warning messages (Figure 1). For optimization of running times, we used cPython package to convert python code in C++ code, which renders faster numerical processing.
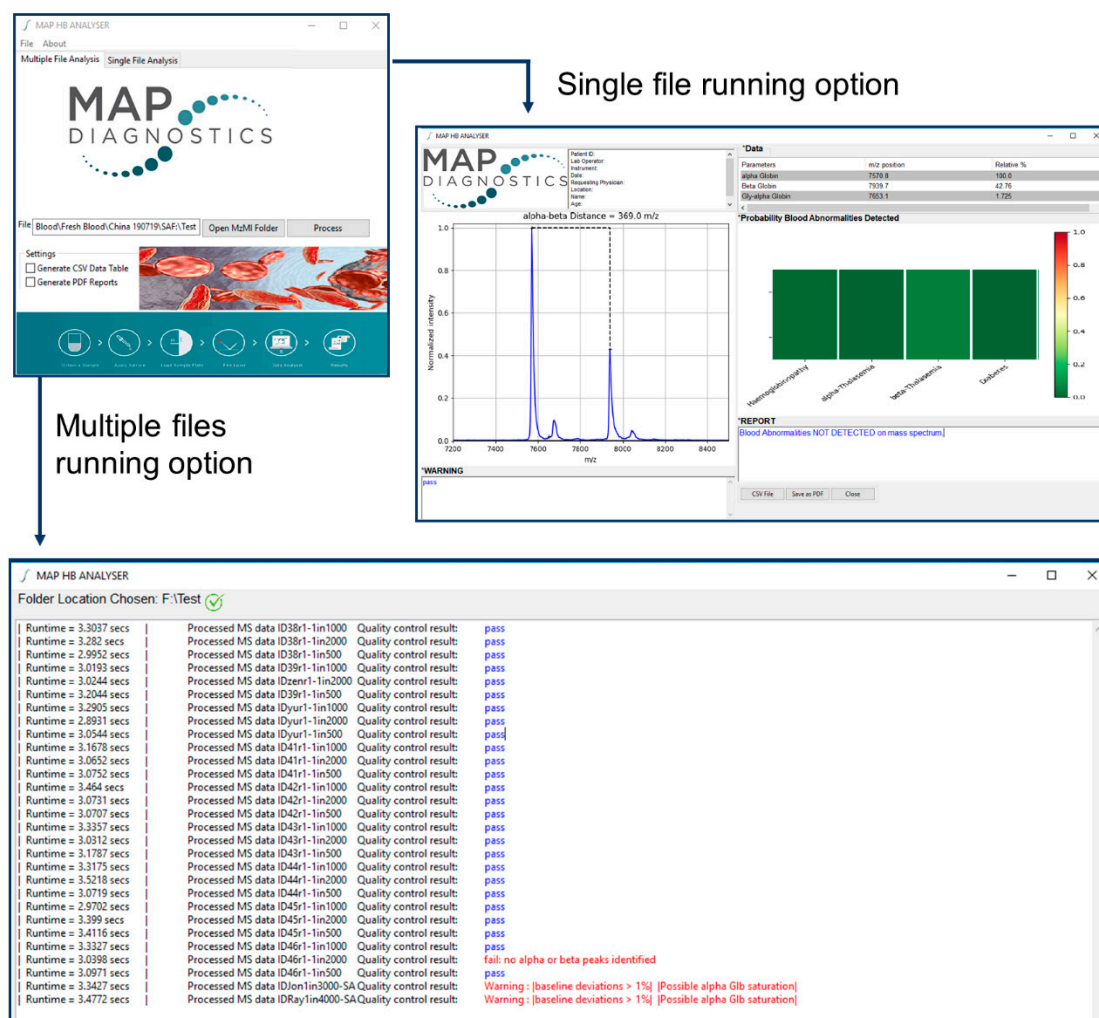

Figure 1. The HbAnalyser software tool version beta 1 for automated detection of blood abnormalities. This version was developed to systematically detect haemoglobinopathies, Thalassemias and diabetes on blood mass spectra from MALDI-ToF instruments.

3. Results

Screening of blood abnormalities was performed on blood samples from 58 individuals using MALDI-ToF technology and a patented methodology [23]. We analyzed the blood samples mass spectra with 3 dilutions to compare data quality, accuracy and robustness of the methodology,

rendering a total of 171 mass spectra. After soaking time of samples, we were able to generate mass spectral data with a high-throughput production rate of 3.4 min, under its maximum sample load capacity (48 samples/plate). We have analyzed the generated data in terms of its quality for using in diagnostics and quantified biomarkers for haemoglobinopathies, thalassemias and diabetes detection on good-quality data. We first performed mass spectral data analysis using a manual approach that requires a fully trained and specialized technician for making a quality control decision and calculating the levels of biomarkers (see Section 2.7). Next, we analyzed the same data with a fully automated approach using software specifically designed for this purpose (see Section 2.8). The results obtained from the automated approach are presented in the following sections of results and compared with a manual approach.

3.1. Efficiency of Automated Approach vs. Manual

To compare the data processing speed of our automated software tool with the manual approach, we have monitored the timings of data analysis for both automated and manual approaches. On average, the observed timing of the fully automated software tool was 3.4 seconds per sample, running under the entire data set (see Supplementary Video). This performance was about 30-fold faster in comparison with the observed times that a trained technician took to process and analyze the same data (~2 min per sample). In general, mass spectral data obtain by MALDI-ToF showed good quality for diagnostic purpose using both manual and automated approaches for quality control assessment (Table 2, see also Supplementary Table S1 in a file for further details). Our data also shows that quality drops slightly with dilution between 1:500 and 1:2000 but still keeping a reasonable acceptance rate. Except for 17 cases, the quality control decision using automated software and manual assessment were in agreement (Table 2). In most cases, spectra had a good quality, where the α-globin and β-globin peaks were identified by both approaches and no spectral issues observed (see example in Figure 2a). For most of the rejected samples, there was not enough signal in the range of α-globin and/or β-globin peaks, showing only noise in the mass spectrum (see example in Figure 2b). For the cases where automated and manual approaches were in disagreement, mass spectra showed lower signal to noise ratios for α-globin and/or β-globin peaks close to quality threshold (see example in Figure 2c).

Table 2. Quality control decision comparison between the automated and manual approach.

Automated	Manual	Dilution 1:500	Dilution 1:1000	Dilution 1:2000
Accepted	Accepted	94.7% (54 cases)	82.5% (47 cases)	66.7% (36 cases)
Rejected	Rejected	3.5% (2 cases)	10.5% (6 cases)	11.1% (6 cases)
Accepted	Rejected	1.8% (1 case)	7.0% (4 cases)	22.2% (12 cases)
Rejected	Accepted	0.0% (0 cases)	0.0% (0 cases)	0.0% (0 cases)

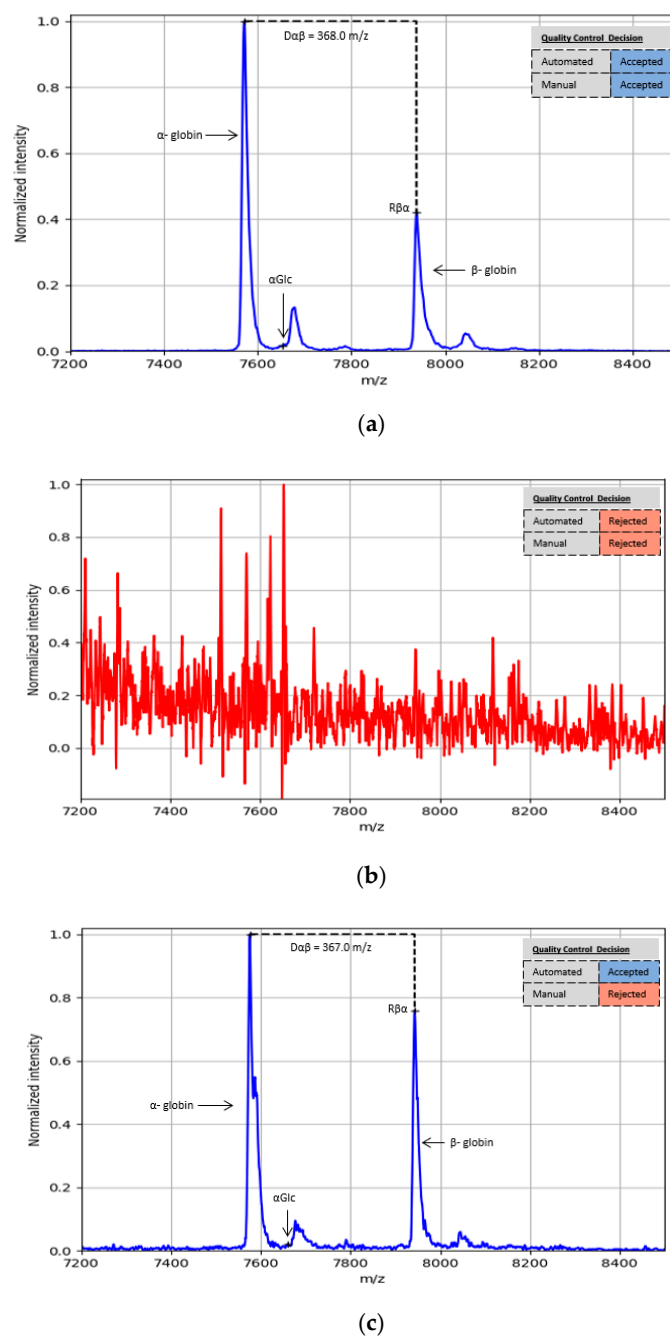(**a**)

(**b**)

(**c**)

Figure 2. Illustrative examples of mass spectra accepted (**a**,**c**) and rejected (**b**,**c**) for clinical diagnostic of blood abnormalities. Quality control decisions made by both manual and automated approaches are depicted in the inset table in grey. The detected peaks of α-globin, β-globin and glycated α-globin are indicated with **+.** The α-globin and β-globin mass distance value (*m/z*) is indicated with black dashed lines and the calculated value indicated below. Mass spectrum figures generated using an automated software tool.

3.2. Variability of Biomarkers

Screening of blood samples of healthy individuals from the UK allowed us to explore the natural variability of biomarkers which is not related to blood abnormalities when applying our MALDI-ToF-based diagnostic technique. We analyzed the variability of biomarkers using both automated software tool and manual approach obtaining consistent results. In comparison with the manual approach, we obtained a 10% ± 15% deviation in each biomarker value, which was mainly due

to rounding errors made by the operator when extracting the peak values in mMass. The biomarkers variability computed using our software tool is presented in Figure 3. For the Dαβ as a biomarker of haemoglobinopathies, we have obtained a natural variability on mass spectra associated with healthy individuals between 366 *m/z* to 371 *m/z* with an average value of 368.2 *m/z* (Figure 3a). This variation can be explained by a slightly drift around both α-globin and β-globin mass peaks discarding any type of variability on the amino acid composition suggestive of genetic mutation. For glycated α-globin ratio (αGlc), the biomarker of diabetes, we obtained a variability that ranges between 0.5% to 3.5% describing 95% of data and with an average value of 1.6% (Figure 3c). Additionally, the variability of these biomarkers shows conservation across different blood dilutions used. This indicates that they are robust markers which are not affected by blood concentration (Figure 3a,c). For the biomarker of thalassemias (Rβα), we obtained ratios of β-globin over α-globin between 20% to 85% with an average value of 51% (Figure 3b). In this case, we obtained an apparent shift in the distribution of the Rβα towards lowering the average value for dilutions of 1:1000 but still keeping similar upper and lower limits in comparison with the dilution of 1:500. Also, the variability of the dilution 1:2000 was lower than the higher dilutions which indicate that dilution may affect the diagnostic power of this biomarker.
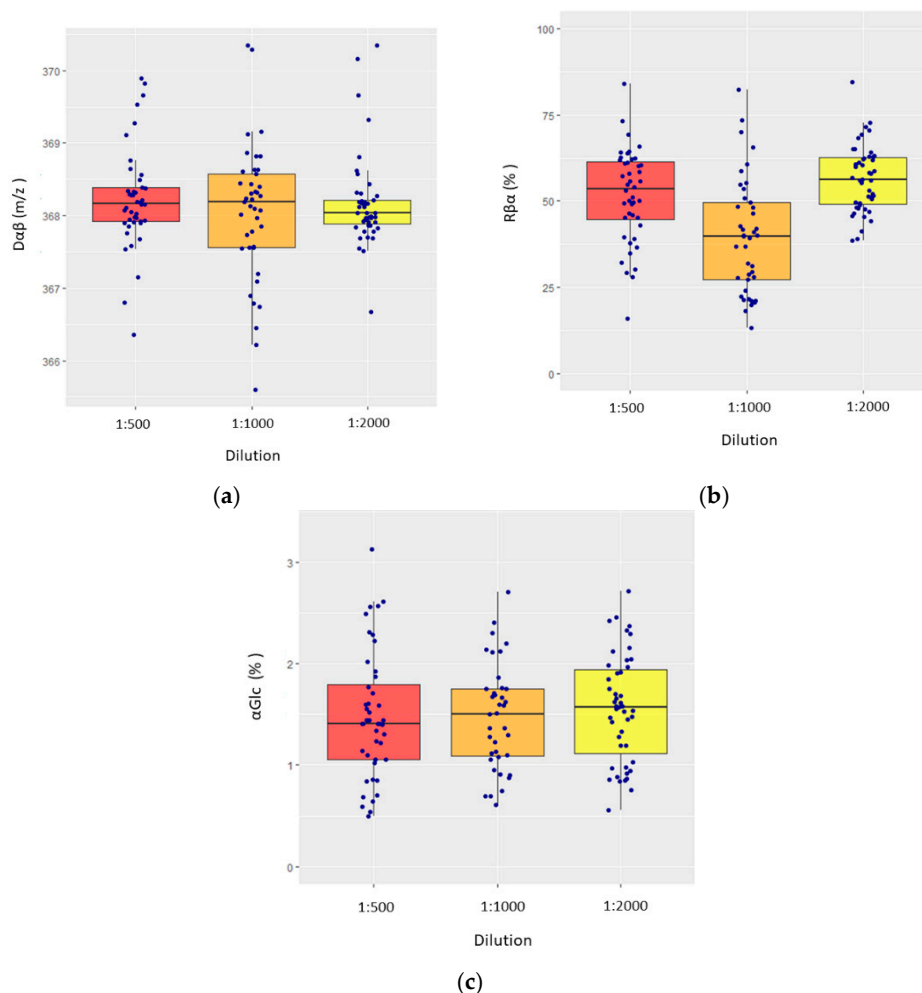

(**a**)　　　　(**b**)

(**c**)

Figure 3. Variability of blood biomarkers in the mass spectra in healthy individuals for three dilutions using the automated approach. (**a**) Variability for the measured mass distance between α-globin and β-globin detected (Dαβ). (**b**) Variability for the calculated ratio between β-globin and α-globin peak intensities (Rβα). (**c**) Variability for the calculated percentage of glycated α-globin peak intensities (αGlc).

3.3. Detected Cases of Blood Abnormalities

In this screening study, we have systematically applied predictive models for haemoglobinopathies, thalassemias and diabetes detection on blood mass spectra using the automated software tool (see Section 2.6 for details). Using these models, we obtained low probabilities ($p < 0.4$) of having each of the analyzed diseases in healthy individuals with blood dilutions of 1:500 and 1:1000. This indicates that the predictive models used were robust in their capacity to discard blood abnormalities accounting for natural variability in these conditions. For the dilution of 1:2000, we also obtained low probabilities ($p < 0.4$) of having diseases in healthy individuals, except for β-thalassemia detection. For this disease, the model predicted low probabilities of having β-thalassemia in 80% of healthy individuals and probabilities between 0.4 and 0.58 for the remaining individuals. This further indicates that lowering the concentration of blood down to 1:2000 increases the uncertainty in predicting β-thalassemia.

During our blood screening study, we have not detected any haemoglobinopathy but detected a possible α-thalassemia case with a probability of 0.68, which has been further analyzed on the mass spectrum (Figure 4). The blood mass spectrum in this particular case shows a reduced abundance of α-globin in comparison with β-globin suggesting α-thalassemia blood abnormality. Despite this abnormality, the other blood markers were within the range of variation of a healthy individual indicating no α-globin saturation. Mass spectrum passed all quality control checking, where no baseline deviations and no presence of contaminants were detected. This further supports that the obtained high β-globin to α-globin ratio is due to an α-thalassemia scenario.
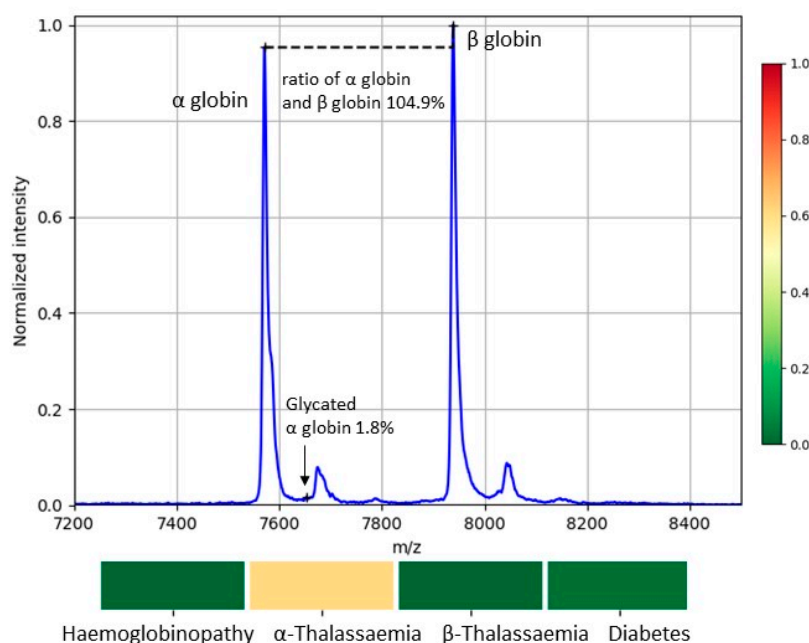

Figure 4. Mass spectrum of the blood of an individual predicted to have α-thalassemia. The detected peaks of α-globin, β-globin and glycated α-globin are indicated with +. The α-globin and β-globin mass distance value (*m/z*) is indicated with black dashed lines and the calculated value indicated below. The mass spectrum was generated with 1:1000 blood dilution and visualized using the automated software tool.

We also detected three potential pre-diabetes cases in our screening, which were discarded from the healthy control group. The mass spectra of these individuals resulted in calculated percentages of glycated α-globin between 4.5% and 7.6% (see example in Figure 5). These percentages were 1–3% higher than those observed for healthy individuals, and near the range of values reported for haemoglobin of pre-diabetic patients (6% to 6.4%) [25]. This further suggests an abnormal percentage of glycated α-globin that can be associated with either a pre-diabetic or controlled diabetic status.
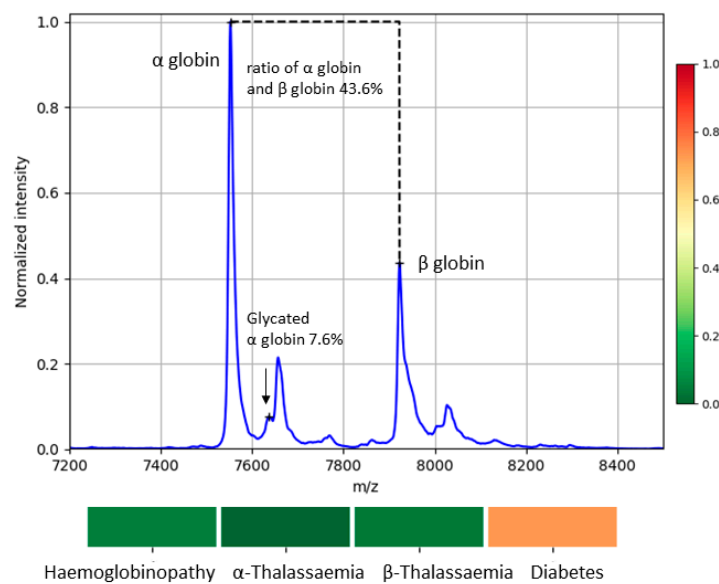
Figure 5. Mass spectrum of a blood sample of a putative pre-diabetic individual. The detected peak of α-globin, β-globin and glycated α-globin are indicated with +. The α-globin and β-globin mass distance value (*m/z*) is indicated with black dashed lines and the calculated value indicated below. The mass spectrum was generated with 1:1000 blood dilution and visualized using the automated software tool.

A sample from a patient with uncontrolled diabetes was also analyzed in this study using the software tool (Figure 6). In this case, the mass spectrum showed a much higher percentage of glycated α-globin in comparison with healthy individuals and putative pre-diabetics. In this case, the glycated α-globin shows a well resolved peak, resulting in a 3-fold higher percentage than the reported threshold for diabetes diagnosis [25]. As expected, the calculated probability using the model was maximum for having diabetes. In general, no other disease was detected in the mass spectra of pre-diabetic and diabetic blood samples and no spectral issues were detected, supporting a correct identification of this disease using our approach.
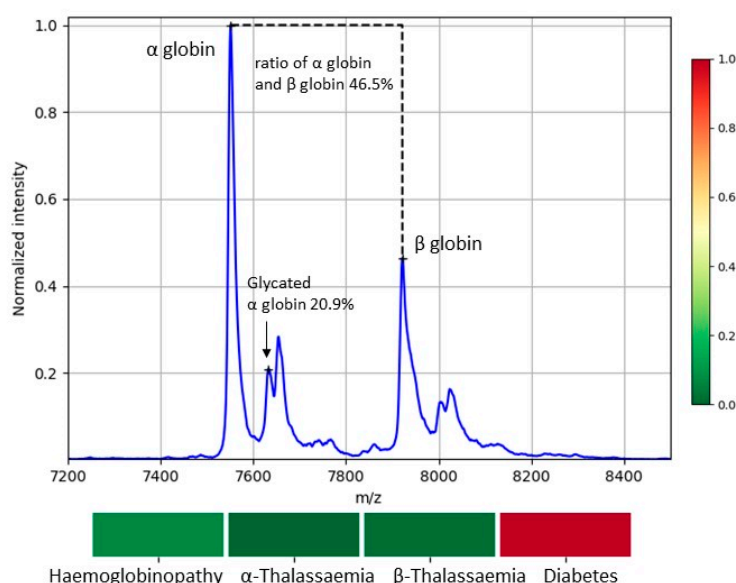
Figure 6. Mass spectrum of a blood sample of an individual with diabetes. The detected peaks of α-globin, β-globin and glycated α-globin are indicated with +. The α-globin and β-globin mass distance value (*m/z*) is indicated with black dashed lines and the calculated value indicated below. The mass spectrum was generated with 1:1000 blood dilution and visualized using the automated software tool.

4. Discussion

In this work, we successfully demonstrated the feasibility of combining MALDI-ToF diagnostic potential with a fully automated software tool towards rapid and affordable screening of blood abnormalities in large populations using the methodology proposed by Iles et al. [18,23]. We demonstrated this application by analyzing a total of 171 mass spectra. The software tool enabled a 30-fold faster analysis, saving human resources while keeping comparable quality control decisions obtained. Here, we have also shown that the software application was able to overcome human bias and rounding errors made by the laboratory technician. While performing manual quality control, no particular cut-off values were taken, leaving it to the experience and decision to the technician, judging the appearance of the spectra was in fact consistent-accurate. On the other hand, the automated quality control is based on the implemented cut off values of the precise threshold which can be further tuned in a consistent way. Thus, we deemed that the automated method is preferable because it performs systematically in the same manner removing the human bias in decision making due to its experience, especially when a peak is too small and may mislead the human eye. This was a major advantage in favor of the automated approach, resulting in higher accuracy in quantifying biomarkers and preventing unnecessary rejection of mass spectra. Together, the software efficiency and accuracy in quality control decision made it possible for the analysis of mass spectral data in detecting blood abnormalities without resorting to highly trained and specialized technicians. This is critical when facing large demand and economic challenges for blood abnormality screening in Asia, Middle East and African countries [1]. Additionally, the rate of systematic processing using the software tool rate was almost instantaneous in comparison with the capacity of data generation by our MALDI-ToF instrument. Thus, we can argue that we have developed a methodological framework that explores the full potential of MALDI-ToF as an ultrafast, high-throughput technique for the detection of blood abnormalities [20].

Blood analysis using MALDI-ToF mass spectrometry of 58 individuals further allowed us to have an idea of the natural variability around the biomarkers for haemoglobinopathies, thalassemias and diabetes detection. This observed variability provides an already useful benchmarking for reference values in terms of biomarkers for clinical diagnostics of haemoglobinopathies, thalassemias, diabetes and even pre-diabetes using mass spectrometry methodology. Moreover, the observed variability

of biomarkers was found to be consistent as a function of blood concentration. It has remarkable importance for clinical diagnostics since haemoglobin concentration in the blood may vary due to dehydration or anaemic state of the patients. However, in terms of generation of predictions, the statistical models of disease used in the tool requires more data from patients with disease to be validated for clinical usage. Yet, our objective was to demonstrate that MALDI-ToF can be used for affordable screening of large populations using a bioinformatics automated system, implemented in a software tool for our clinical laboratory. Despite our modelling assumptions and limitations, the used models have still made accurate predictions which we considered an illustration of the future potential of the tool when more accurate models are available.

It is important to mention that we developed and optimized this tool for the Shimadzu system model H8020 under our laboratory conditions. For other MALDI-ToF models and laboratory experimental settings, we cannot guaranty that the present tool works with properly. This is mainly because different models may use different output file systems and we have optimized the quality control decision algorithm and biomarker detection algorithm for our experimental settings. Nevertheless, we have previously worked with other models and tuned the optimization parameters to keep a good performance of the method, which allow the adaptation of the tool to other laboratory systems in future versions.

Based on the obtained variability in biomarkers and empirical knowledge related to disease detection, we were able to successfully apply predictive models that recapitulated healthy individuals and detected some cases of blood abnormalities. Although these models still require more data with confirmed disease cases to be validated for its diagnostic power, our results already provide a demonstration of their potential; in particular, they may help future clinicians on relating the quantification of biomarkers with the likelihood of having a particular disease based on populational data. This would be further advantageous in future clinics when making correlations based on the variability of their particular population instead of published studies on other populations. Thus, future studies and the accumulation of clinical data would be fundamental for improving and validating such predictive models. For this purpose, the implementation of machine learning algorithms as the ones already applied for mass spectral data would likely boost the accuracy of our predictive models [28].

Taken together, the software tool presented shows remarkable potential as a future tool in clinical diagnostics. Even though the tool was developed on a Shimadzu system and calibrated for our laboratory, it can be adapted to other MALDI-ToF models and laboratories in future versions. Additionally, the tool requires further improvements towards clinical implementation. Thus, we are including the detection of more blood disorders such as sickle cell disease, blood poisoning, effects of air pollution and even the detection of specific variants of haemoglobinopathies and thalassemias based on already developed methodologies [18,23,29]. This will substantially increase the multiplicity of the tool for blood abnormalities and make it even more applicable. Theoretically, the implementation of these features is feasible and is considered to be included in the future extension of the tool. Software improvements towards adding patient history tracking and monitoring would also be a key feature for clinics, especially in the case of diabetes and pre-diabetes control. Finally, we are also implementing a systematic and custom-made report, which would be handy for communication between laboratory, doctor and a patient.

5. Conclusions

In this work, we proved that rapid and affordable screening of multiple genetic and metabolic diseases on large populations on a single blood sample is possible using MALDI-ToF mass spectra spectrometry. Here, we demonstrated that this technological breakthrough is made by using automated quality control assessment and state-of-the-art blood mass spectrum analysis. Moreover, we showed a novel software that shows promising application as a future clinical diagnostic tool for laboratory detection of haemoglobinopathies, thalassemias and diabetes in blood samples.

6. Patents

Method for detection of blood abnormalities in blood using mass spectrometry is under the Patent: PCT/GB2015/052491 WO/2016/030688 2016.

References

1. De Benoist, B.; McLean, E.; Egli, I.; Cogswell, M. Worldwide prevalence of anaemia 1993-2005: WHO Global Database on Anaemia. *WHO Glob. Database Anaemia* **2008**, *1*, 1–51.

2. *Secretariat Sickle-Cell Anaemia*; Report by the Secretariat Prevalence of Sickle-Cell Anaemia; World Health Organization: Geneva, Switzerland, 2006; pp. 1–5.

3. Basran, R.K.; Patterson, M.; Walker, L.; Nakamura, L.M.; Eng, B.; Chui, D.H.; Waye, J.S. Prenatal Diagnosis of Hemoglobinopathies in Ontario, Canada. *Ann. N. Y. Acad. Sci.* **2005**, *1054*, 507–510. [CrossRef] [PubMed]

4. AlSwaidi, F.M.; O'Brien, S.J. Premarital screening programmes for haemoglobinopathies, HIV and hepatitis viruses: Review and factors affecting their success. *J. Med. Screen.* **2009**, *16*, 22–28. [CrossRef] [PubMed]

5. Ford, J.C.; Chitayat, D.; Désilets, V.A.; Farrell, S.A.; Geraghty, M.; Nelson, T.; Nikkel, S.M.; Shugar, A.; Skidmore, D.; Allen, V.M.; et al. Carrier screening for thalassemia and hemoglobinopathies in Canada. *J. Obstet. Gynaecol. Can.* **2008**, *30*, 950–959.

6. Bertram, M.Y.; Vos, T. Quantifying the duration of pre-diabetes. *Aust. N. Z. J. Public Health* **2010**, *34*, 311–314. [CrossRef]

7. Choi, S.B.; Kim, W.J.; Yoo, T.K.; Park, J.S.; Chung, J.W.; Lee, Y.-H.; Kang, E.S.; Kim, D.W. Screening for Prediabetes Using Machine Learning Models. *Comput. Math. Methods Med.* **2014**, *2014*, 618976. [CrossRef]

8. Cefalu, W.T. "Prediabetes": Are There Problems with This Label? No, We Need Heightened Awareness of This Condition! *Diabetes Care* **2016**, *39*, 1472–1477. [CrossRef]

9. Goonasekera, H.; Paththinige, C.; Dissanayake, V. Population Screening for Hemoglobinopathies. *Annu. Rev. Genom. Hum. Genet.* **2018**, *19*, 355–380. [CrossRef]

10. Angeletti, S. Matrix assisted laser desorption time of flight mass spectrometry (MALDI-TOF MS) in clinical microbiology. *J. Microbiol. Methods* **2017**, *138*, 20–29. [CrossRef]

11. Seng, P.; Drancourt, M.; Gouriet, F.; La Scola, B.; Fournier, P.; Rolain, J.M.; Raoult, D. Ongoing Revolution in Bacteriology: Routine Identification of Bacteria by Matrix-Assisted Laser Desorption Ionization Time-of-Flight Mass Spectrometry. *Clin. Infect. Dis.* **2009**, *49*, 543–551. [CrossRef]

12. Nomura, F. Proteome-based bacterial identification using matrix-assisted laser desorption ionization–time of flight mass spectrometry (MALDI-TOF MS): A revolutionary shift in clinical diagnostic microbiology. *Biochim. et Biophys. Acta (BBA)-Proteins Proteom.* **2015**, *1854*, 528–537. [CrossRef] [PubMed]

13. Gaillot, O.; Blondiaux, N.; Loïez, C.; Wallet, F.; Lemaître, N.; Herwegh, S.; Courcol, R.J. Cost-Effectiveness of Switch to Matrix-Assisted Laser Desorption Ionization-Time of Flight Mass Spectrometry for Routine Bacterial Identification. *J. Clin. Microbiol.* **2011**, *49*, 4412. [CrossRef] [PubMed]

14. Tan, K.E.; Ellis, B.C.; Lee, R.; Stamper, P.D.; Zhang, S.X.; Carroll, K.C. Prospective Evaluation of a Matrix-Assisted Laser Desorption Ionization-Time of Flight Mass Spectrometry System in a Hospital Clinical Microbiology Laboratory for Identification of Bacteria and Yeasts: A Bench-by-Bench Study for Assessing the Impact on Time to Identification and Cost-Effectiveness. *J. Clin. Microbiol.* **2012**, *50*, 3301–3308. [PubMed]

15. Duncan, M.W.; Nedelkov, D.; Walsh, R.; Hattan, S.J. Applications of MALDI Mass Spectrometry in Clinical Chemistry. *Clin. Chem.* **2015**, *62*, 134–143. [CrossRef]

16. Iles, R.K.; Cole, L.A.; Butler, S.A. Direct Analysis of hCGβcf Glycosylation in Normal and Aberrant Pregnancy by Matrix-Assisted Laser Desorption/Ionization Time-of-Flight Mass Spectrometry. *Int. J. Mol. Sci.* **2014**, *15*, 10067–10082. [CrossRef]

17. Biroccio, A.; Urbani, A.; Massoud, R.; Di Ilio, C.; Sacchetta, P.; Bernardini, S.; Cortese, C.; Federici, G. A quantitative method for the analysis of glycated and glutathionylated hemoglobin by matrix-assisted laser desorption ionization-time of flight mass spectrometry. *Anal. Biochem.* **2005**, *336*, 279–288. [CrossRef]

18. Iles, R.; Mahmoud, N. Future Laboratory Medicine: Rapid, Efficient and Affordable Screen-ing for Haemoglobinopathies by MALDI-ToF Mass Spectrometry. *Adv. Biochem. Biotehcnol.* **2018**, *2017*, 1–5.

19. Butler, S.A.; Luttoo, J.; Freire, M.O.T.; Abban, T.K.; Borrelli, P.T.A.; Iles, R.K. Human chorionic gonadotropin (hCG) in the secretome of cultured embryos: Hyperglycosylated hCG and hCG-free beta subunit are potential markers for infertility management and treatment. *Reprod. Sci.* **2013**, *20*, 1038–1045. [CrossRef]

20. Haslam, C.; Hellicar, J.; Dunn, A.; Fuetterer, A.; Hardy, N.; Marshall, P.; Paape, R.; Pemberton, M.; Resemannand, A.; Leveridge, M. The Evolution of MALDI-TOF Mass Spectrometry toward Ultra-High-Throughput Screening: 1536-Well Format and Beyond. *J. Biomol. Screen.* **2016**, *21*, 176–186. [CrossRef]

21. Pan, S.; Zhang, H.; Rush, J.; Eng, J.; Zhang, N.; Patterson, D.; Comb, M.J.; Aebersold, R. High Throughput Proteome Screening for Biomarker Detection. *Mol. Cell. Proteom.* **2005**, *4*, 182–190. [CrossRef]

22. Pais, R.J.; Zmuidinaite, R.; Butler, S.; Iles, R. An automated workflow for MALDI-ToF mass spectra pattern identification on large data sets: An application to detect aneuploidies from pregnancy urine. *Inform. Med. Unlocked* **2019**, *16*, 100194. [CrossRef]

23. Iles, R.K.; Iles, J.K.; Abban, T.; Docherty, S.M.; Nasse, M. Method for Detecting Abnormalities in Hemoglobin. WO Patent 2016/030688, July 2019.

24. Consultation, W. *Use of Glycated Haemoglobin (HbA1c) in the Diagnosis of Diabetes Mellitus*; World Health Organization: Geneva, Switzerland, 2011.

25. Punthakee, Z.; Goldenberg, R.; Katz, P. Definition, Classification and Diagnosis of Diabetes, Prediabetes and Metabolic Syndrome. *Can. J. Diabetes* **2018**, *42*, S10–S15. [CrossRef] [PubMed]

26. Strohalm, M.; Hassman, M.; Kosata, B.; Kodíček, M. mMass data miner: An open source alternative for mass spectrometric data analysis. *Rapid Commun. Mass Spectrom.* **2008**, *22*, 905–908. [CrossRef]

27. Strohalm, M.; Kavan, D.; Novák, P.; Volný, M.; Havlíček, V. mMass3: A Cross-Platform Software Environment for Precise Analysis of Mass Spectrometric Data. *Anal. Chem.* **2010**, *82*, 4648–4651. [CrossRef]

28. Swan, A.L.; Mobasheri, A.; Allaway, D.; Liddell, S.; Bacardit, J. Application of machine learning to proteomics data: Classification and biomarker identification in postgenomics biology. *OMICS J. Integr. Boil.* **2013**, *17*, 595–610. [CrossRef] [PubMed]

29. Greene, D.N.; Vaugn, C.P.; Crews, B.O.; Agarwal, A.M. Advances in detection of hemoglobinopathies. *Clin. Chim. Acta* **2014**, *439*, 50–57. [CrossRef] [PubMed]