



## ARTICLE



<https://doi.org/10.1038/s42003-020-01262-z>

OPEN

# Machine learning assistive rapid, label-free molecular phenotyping of blood with two-dimensional NMR correlational spectroscopy

Weng Kung Peng <sup>1✉</sup>, Tian-Tsong Ng<sup>2</sup> & Tze Ping Loh <sup>3✉</sup>

Translation of the findings in basic science and clinical research into routine practice is hampered by large variations in human phenotype. Developments in genotyping and phenotyping, such as proteomics and lipidomics, are beginning to address these limitations. In this work, we developed a new methodology for rapid, label-free molecular phenotyping of biological fluids (e.g., blood) by exploiting the recent advances in fast and highly efficient multidimensional inverse Laplace decomposition technique. We demonstrated that using two-dimensional  $T_1$ - $T_2$  correlational spectroscopy on a single drop of blood ( $<5 \mu\text{L}$ ), a highly time- and patient-specific 'molecular fingerprint' can be obtained in minutes. Machine learning techniques were introduced to transform the NMR correlational map into user-friendly information for point-of-care disease diagnostic and monitoring. The clinical utilities of this technique were demonstrated through the direct analysis of human whole blood in various physiological (e.g., oxygenated/deoxygenated states) and pathological (e.g., blood oxidation, hemoglobinopathies) conditions.

<sup>1</sup>Precision Medicine – Engineering Group, International Iberian Nanotechnology Laboratory, 4715 330 Braga, Portugal. <sup>2</sup>Institute for Infocomm Research, Fusionopolis Way, Singapore, Singapore. <sup>3</sup>Department of Laboratory Medicine, National University Hospital, Singapore, Singapore. ✉email: [weng.kung@inl.int](mailto:weng.kung@inl.int); [tze\\_ping\\_loh@nuhs.edu.sg](mailto:tze_ping_loh@nuhs.edu.sg)

High-resolution nuclear magnetic resonance (NMR) spectroscopy is a powerful and attractive technique in biochemistry (e.g., for structural protein analysis<sup>1</sup>, characterizing metabolomics responses in biological samples<sup>2–4</sup>) and inorganic chemistry<sup>5</sup>. However, high-resolution NMR systems are large, expensive and incompatible with in situ or portable applications. There is an increasing demand for low-field portable NMR system for use in food sciences<sup>6</sup>, oil-gas exploration<sup>7</sup>, and clinical diagnostic at point-of-care testing (POCT)<sup>8–11</sup>. In high-field NMR, biochemical information is typically detected and encoded in the frequency domain (“chemical shift”), in which the spectral resolution scale with respect to the external magnetic field. This reduces its portability and limit its downstream application in a large scale manner.

However, biochemical and biophysical information (e.g., molecular rotational, diffusional motion) can also be encoded in the relaxation times frame, namely the longitudinal ( $T_1$ ) and transverse ( $T_2$ ) using NMR-based POCT. In addition, molecular information in the time-domain can be inversely decoded with the availability of fast and reliable Laplace inversion algorithm<sup>7,12</sup>. This can provide parallel information that is not available in the traditional NMR frequency domain based spectra.

In recent years, significant advances in NMR system miniaturization<sup>8,13–15</sup> (e.g., electronic console<sup>13,14,16,17</sup>, radio-frequency probe<sup>9,10,18–20</sup>, microfluidic-based chip<sup>21,22</sup>) utilizing small foot-print permanent magnetic (<1 Tesla) for one-dimensional NMR relaxometry on water-proton (e.g.,  $T_2$ -relaxation) have been widely applied for point-of-care medical testing<sup>8,9,14</sup>. These include immuno-magnetic labeled (e.g., tumour cells<sup>8,22</sup>, tuberculosis<sup>23</sup> and magneto-DNA detection of bacteria<sup>24</sup>) and the label-free detection of various pathological states such as oxygenation<sup>20</sup>/oxidation level<sup>10</sup> of the blood, malaria screening<sup>9,25</sup>, and rapid phenotyping of oxidative stress in diabetes mellitus<sup>26,27</sup>.

We demonstrated (to the best of our knowledge) the first unique two-dimensional ‘molecular fingerprint’ of a single drop of blood (<5  $\mu$ L) obtained in minutes using two dimensional  $T_1$ - $T_2$  correlational spectroscopy with an inexpensive, benchtop-sized NMR spectrometer<sup>28,29</sup>. By exploiting the recent development of fast and highly efficient multi-dimensional inverse Laplace decomposition algorithm<sup>7,30</sup>, unique two-dimensional signature of various hemoglobin (Hb) derivatives with respect to its magnetic resonance relaxation reservoirs in oxygenated (oxy-Hb), deoxygenated (deoxy-Hb) and oxidized (oxidized Hb) states were observed for the first time (to the best of our knowledge) and its phenotypic expression in various pathological states (e.g., blood oxidation, hemoglobinopathies) are reported in this work. Machine-learning techniques (e.g., multidimensional scaling (MDS), t-SNE, Isomap) were introduced to transform the NMR correlational maps into user-friendly information for medical decision making. We report that the supervised models (e.g., neural network) were at least on par or outperformed the average trained human being in performing the deep image analysis of molecular fingerprint of red blood cells (RBCs).

## Results

**Water-protein interactions in blood microenvironment.** Freshly collected whole blood samples containing predominantly the oxy-Hb were collected from healthy donors (‘wild-type’). Oxygenation and re-oxygenation was achieved with rigorous pipetting in ambient air. Using microcapillary tube, the whole blood was sampled and spun (6000  $\times$  g, 1 min) into narrowband of RBCs for micro NMR measurements (Figs. 1–2).

Three peaks (*R*-peak, *S*-peak and *T*-peak) with ( $T_2 = 141$  ms,  $T_1 = 562$  ms), ( $T_2 = 4.47$  ms,  $T_1 = 335$  ms) and ( $T_2 = 1.12$  ms,

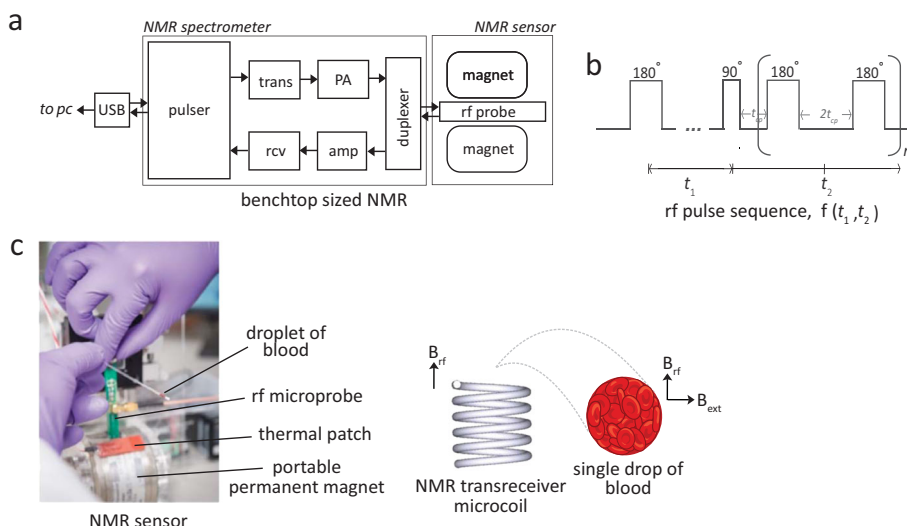
$T_1 = 188$  ms) respectively were observed from the  $T_1$ - $T_2$  correlational spectroscopy performed on the water-proton nuclei ( $^1\text{H}$ ) of the RBCs (Fig. 3a). It appeared that RBCs microenvironment could be decomposed into two major relaxation reservoirs, consisting of one slow relaxation component (*R*-peak), and two fast relaxation components (*S*-peak, *T*-peak), attributed to the interaction of the water molecules with its’ respective micro-environment i.e., bulk water, intermediate hydration layer, macromolecules protein, respectively (Supplementary Fig. 1). Water molecules are subjected to diverse dynamic processes as a result of their interaction with variety of sites/functional groups.

The significantly large signal intensity (and slowest relaxation component) of *R*-peak is attributable to bulk water molecules which makes up more than 98% of the total mass-ratio of RBCs. The bulk water has minimal and indirect contact with macromolecules protein (through long-range dipolar couplings), and hence the weakest water-protein interactions. The relaxation dephasing system came predominantly from the dipole-dipole homonuclei coupling of water-to-water network. On the other hand, the presence of two distinct individual peaks (i.e., *S*-peak, *T*-peak) suggested that the fast relaxation component can be further resolved into sub-regions<sup>31,32</sup>. The *S*-peak is the water molecules at the intermediate hydration layer, and the *T*-peak are water molecules, which came into direct contact with the surface of macromolecular protein. Dortch et al. and McDonald et al., proposed the idea of exchange peaks<sup>33</sup> and surface relaxation<sup>34</sup>, respectively, but the observation in this work is in consistent with the three peaks model proposed by Lores et al. and Thompson et al.<sup>35,36</sup>.

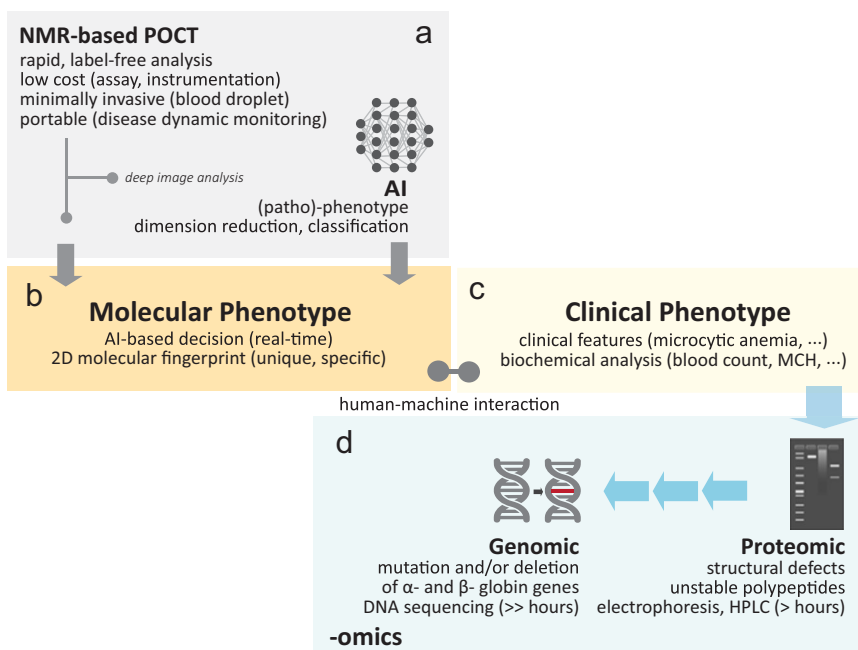
Interestingly, each peak (*R*-peak, *S*-peak, *T*-peak) possess consistent and yet unique ratio of  $T_1/T_2$  of (3.99, 74.90, 167.86), respectively, which appeared to characterize the degree of water-protein interactions (Fig. 3b and Table 1). We define here the  $T_1/T_2$  ratio as A-ratio. With increased water-protein interactions, the motion of water-proton was drastically slower and restricted (and hence the reduced  $T_1$  relaxation and  $T_2$  relaxation). The spin-spin relaxation appeared to be much more efficient (shorter  $T_2$  relaxation) relative to its’ spin-lattice relaxation counterpart and hence a large A-ratio. In contrast, an unbound free molecules in the extreme fast motion region, possess large  $T_1$  relaxation and  $T_2$  relaxation, with A-ratio approaches unity ( $\sim 1$ ). Importantly, the relaxation profile forms unique and specific two-dimensional ‘molecular fingerprint’ of each individual that is very sensitive to its’ molecular microenvironment measurable at the timescales of NMR relaxation times.

**Oxidative degradation of hemoglobin in blood.** Freshly collected whole blood sample which consists of predominantly the oxy-Hb was oxidized to oxidized Hb in the presence of sodium nitrite, and spun down for NMR measurements (see Methods). The relaxation times of the three major peaks were (*R*-peak:  $T_2 = 120$  ms,  $T_1 = 217$  ms), (*S*-peak:  $T_2 = 4.18$  ms,  $T_1 = 120$  ms), and (*T*-peak:  $T_2 = 1.34$  ms,  $T_1 = 50.3$  ms) in oxidized state reduced considerably as compared to the baseline oxygenated state (non-oxidized, diamagnetic state) (Fig. 4a, b). The presence of excessive oxidized Hb in blood causes serious tissue hypoxia, a pathological state known clinically as methemoglobinemia<sup>37</sup>.

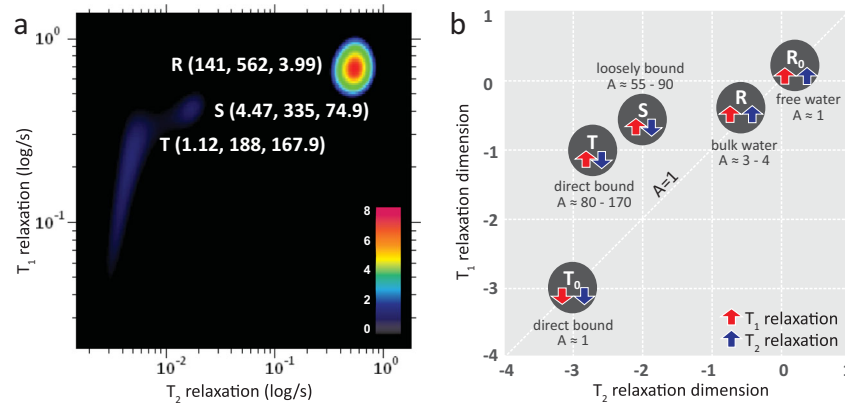
The marked relaxation enhancement observed was due to the presence of five unpaired electrons in the ferric iron ( $\text{Fe}^{3+}$ ), which acted as the paramagnetic relaxation center<sup>37,38</sup>. The magnetic moment of ferric iron is 1000-fold higher than that of one single proton<sup>37,39</sup>. Significantly, due to the long-range dipolar nuclei-electron, the paramagnetism of the unpaired electrons had considerable effect on the bulk water molecules (*R*-peak). In contrast to the oxygenated states (in diamagnetic state), the spin-lattice relaxation



**Fig. 1 Two-dimensional NMR  $T_1$ - $T_2$  correlational spectroscopy for molecular phenotyping of blood.** **a** Schematic diagram of the bench-top sized NMR-based POCT system. The applied radio frequencies were centered at 21.57 MHz, which corresponds to the Larmor frequency of water-proton in 0.5 Tesla of the permanent magnet. The 90-degree pulse used is 10  $\mu$ s. The whole system is lightweight (<2 kg) and portable suitable for in situ measurements. The abbreviations are; USB Universal Serial Bus, trans Transmitter, rcv Receiver, amp pre-amplifier, PA power amplifier, rf radio-frequency, and PC personal computer. **b** The pulse sequence used for the  $T_1$ - $T_2$  correlational spectroscopy is the modified inversion recovery with CPMG observation. It is encoded for a period of  $t_1$  and subsequently spaced for a period of  $t_2$  for n-train pulses, in entirely in analogous to the two-dimensional NMR spectroscopy in the frequency domain. The relaxation properties can be used as a highly sensitive and specific molecular probe, and provide important molecular motion (e.g., correlational relaxation, diffusion properties), which is not readily available in NMR spectra in the frequency domain. **c** A single drop of whole blood contained in a microcapillary tube was spun using standard hematocrit centrifuge (6000  $\times$  g, 1 min) to separate and concentrate the RBCs from the plasma. The capillary tube is then loaded into a permanent magnet. The tube was adjusted as such that the radio frequency coil (inner diameter of 1.20 mm) focuses on the packed RBCs (enrichment part). This is essential to have ‘clean signal’ from the RBCs without (or with minimal) interference of blood plasma.



**Fig. 2 A proposed scheme of human-machine interaction for rapid, label-free disease detection in clinical hemoglobinopathies.** **a** The NMR-based POCT is used with (or without) the assistant of artificial intelligence (AI). **b** The highly unique and detailed 2D magnetic resonance-based molecular fingerprint can be used directly (without AI) for rapid screening. **c** Clinical phenotype (e.g., clinical representation) can be bias due to subjective human judgment. With AI, deep image analysis (e.g., hierarchical clustering, dimension reduction) were performed to transform the highly complicated data (e.g., hyper dimension) into human friendly information to assist in medical decision making (e.g., diagnostic, staging) in real-time mode (Fig. 6). **d** Multi-omics information (e.g., proteomics, genomics) may be performed simultaneously to confirm the genetic variants and/or other anomalies. Back-end laboratory and time consuming test (e.g., high-performance liquid chromatography (HPLC)) may be by-passed depending on the outcome of the molecular phenotyping.



**Fig. 3 The  $T_1$ - $T_2$  correlational spectrum of blood microenvironment.** **a** The decomposed relaxation reservoirs ( $R$ -peak,  $S$ -peak, and  $T$ -peak) of packed red blood cells microenvironment with the hemoglobin in oxygenated state. The coordinate is represented as ( $T_2$  relaxation (in ms),  $T_1$  relaxation (in ms),  $A$ -ratio (unitless)).  $A$ -ratio is the ratio between  $T_1/T_2$ . **b** The multiple relaxation reservoirs of the blood microenvironment in the  $T_1$ - $T_2$  correlational spectrum in log-log plot; i.e., the bulk water ( $R$ -peak), hydration layer ( $S$ -peak), and direct macromolecular protein interaction ( $T$ -peak) for hemoglobin in oxygenated state. In the oxidized state, the  $T$ -peak dropped substantially ( $T_0$ -peak). The unbound molecule,  $R_0$  (e.g., free water) located on the diagonal line ( $A$ -ratio approaches unity).

**Table 1 The decomposed relaxation reservoirs of packed red blood cells with the hemoglobin in (a) oxygenated, (b) oxidized, and (c) deoxygenated states.**

Bulk water, $R$ -peak	$T_1$ (ms)	$T_2$ (ms)	$A$ -ratio
a. oxygenated state	562	141	3.99
b. oxidized state	217	120	1.81
c. deoxygenated state	463	102	4.53
Hydration layer, $S$ -peak	$T_1$ (ms)	$T_2$ (ms)	$A$ -ratio
a. oxygenated state	335	4.47	74.94
b. oxidized state	120	4.18	28.71
c. deoxygenated state	242	2.71	89.30
Direct bound macromolecules, $T$ -peak	$T_1$ (ms)	$T_2$ (ms)	$A$ -ratio
a. oxygenated state	188	1.12	167.86
b. oxidized state, ( $T_i$ )	50.3	1.34	37.54
b. oxidized state, ( $T_0$ )	2.43	0.78	3.12
c. deoxygenated state	175	0.565	309.73

$A$ -ratio is the ratio between  $T_1/T_2$ .

effect in oxidized states (in paramagnetic state) appeared to be much more efficient in comparison to the spin-spin relaxation effect and hence the reduction in  $A$ -ratio = 1.81 (Table 1).

A distinctively long stretch of  $T_1$ -relaxation distribution, extending across two orders of magnitude (ca., 1 ms to 100 ms along the  $T_1$  dimension) displayed by the protein-bound water-proton molecules (from  $T$ -peak to  $T_0$ -peak). The ‘relaxation tail’ originating from ( $T_2 = 1.34$  ms,  $T_1 = 50.3$  ms) to ( $T_2 = 0.78$  ms,  $T_1 = 2.43$  ms), notably became a distinctive feature of oxidized Hb. This is due to the distance ( $r$ )-dependent paramagnetism effect, in which the relaxation efficiency reduced at the rate of  $1/r^6$  from its relaxation center<sup>37</sup>. As the proton nuclei approach the relaxation center (of the unpaired electron), the  $T_1$ - and  $T_2$ -relaxation components reduced to a comparable rate ( $A$ -ratio approaching unity,  $T_0$  in Fig. 3b). The gradual process of Hb oxidation under the exposure of mild oxidant were captured in a well-controlled manner confirmed the existence of transitional states in the formation of ‘tail’ (Supplementary Fig. 2).

On the other hand, the protein-bound water molecules ( $T$ -peak) in the deoxygenated states, exhibited profound  $T_2$  shortening

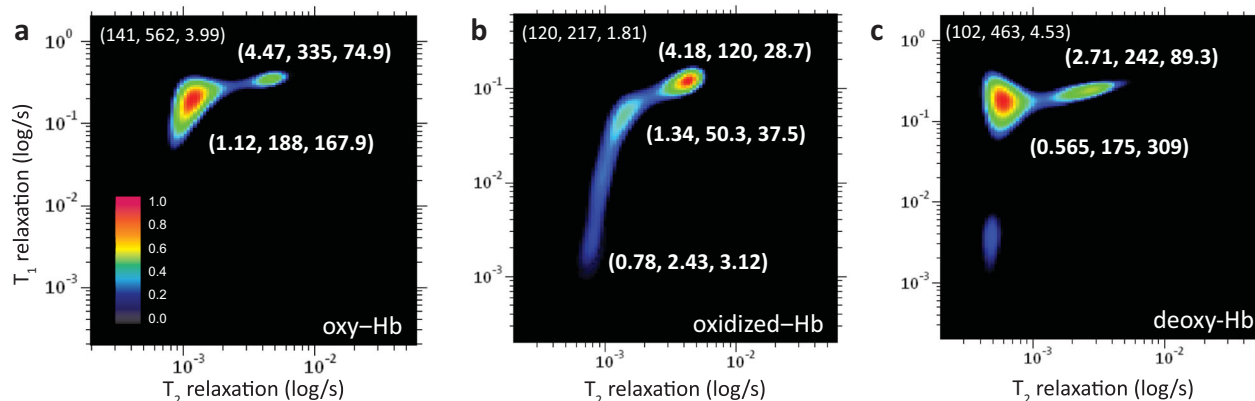
(0.565 ms) with relatively very little  $T_1$  shortening (175 ms) due to the short relaxation time of electron and its obscure protein configuration<sup>40</sup>. As a result, the  $A$ -ratio of deoxy-Hb (309) is distinctively larger than its oxy-Hb (167.9) and oxidized Hb (37.5) counterparts (Fig. 4c).

#### Rapid molecular phenotyping in clinical hemoglobinopathies.

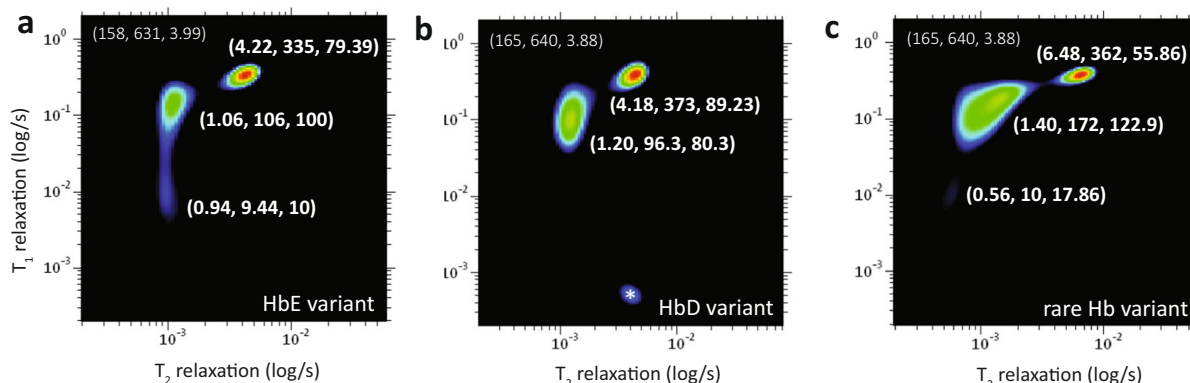
We demonstrated the clinical utility of molecular phenotyping in clinical hemoglobinopathies by mapping out the spectrum of heterozygous HbE, HbD and a heterozygous beta thalassemia (HBB:c.27\_28insG) variants (Fig. 5 and Table 2). An additional six other Hb variants (in Supplementary Fig. 3b) were received for machine learning and blind test studies (Table 3 and Fig. 6). A limitation of this study was that the current study only involve heterozygous HbE phenotype. Given the low prevalence of homozygous HbE variant phenotype (~0.1%) in our population<sup>41</sup>, therefore, we were unable to include such subject during the study period. The Hb variants were first identified by a cation-exchange high-performance liquid chromatography method (Bio-Rad Variant II analyzer) and further confirmed by capillary electrophoresis (Sebia CAPILLARYS 2 analyzer) and genotyping. NMR measurements were carried out in its native state (without any chemical treatment) of the spun down packed RBCs.

The Hb genotyping identified single nucleotide polymorphism in the  $\beta$ -globin in the first and second samples, which was consistent with HbE (Fig. 5a) and HbD variant (Fig. 5b). A third rare Hb variant samples were identified with a G insertion at codon 27 of the  $\beta$ -globin gene (Fig. 5c). These hemoglobin variants exhibit similar clinical phenotype such as mild hemolysis and susceptible to oxidation<sup>42,43</sup>. The two-dimensional correlational mapping of Hb variants (Fig. 5a-c) revealed an unusual spectrum characteristic as compared to wild-type RBCs (Supplementary Fig. 3). The HbE variant ( $T_2 = 1.06$  ms,  $T_1 = 106$  ms), HbD variant ( $T_2 = 1.20$  ms,  $T_1 = 96.3$  ms), and the beta thalassemia variant ( $T_2 = 1.40$  ms,  $T_1 = 172$  ms) appears to have large and distorted  $T$ -peak with relatively short  $T_1$ - and  $T_2$ - relaxations as compared to wild-type Hb ( $T_2 = 1.12$  ms,  $T_1 = 188$  ms). The  $T$ -peak dispersion for the beta thalassemia variant with a mutated  $\beta$ -globin chain was particularly large with a flat plateau, suggesting that frame shift mutation causes a greater amount of hemoglobin instability<sup>42</sup> (Fig. 5c).

In addition, the Hb variants appear to have much higher concentration of oxidized Hb as compared to the wild-type (Supplementary Fig. 3a).  $T_1$ -relaxation stretching was observed for



**Fig. 4** The  $T_1$ - $T_2$  correlational spectrum of blood microenvironment of (wild type) packed red blood cells. There were in various physiological states; **a** oxygenated, **b** oxidized, and **c** deoxygenated states. The zoom-in details of decomposed relaxation reservoirs for fast relaxation components ( $S$ -peak and  $T$ -peak) and the slow relaxation component (bulk water molecules,  $R$ -peak) is not shown. The coordinate for  $R$ -peak is indicated at upper left of the spectrum. The coordinate is represented by ( $T_2$  relaxation (in ms),  $T_1$  relaxation (in ms),  $A$ -ratio). Freshly prepared oxy-Hb was subjected to oxidation with 10 mM sodium nitrite for 45 min, and sodium dithionite (in excess) for 40 min to chemically locked the Hb in the deoxygenated state. All the samples were washed thrice and resuspended into 1x PBS for micro MR measurements. The experimental parameters used were echo time = 200  $\mu$ s,  $T_1$ -incremental steps = 32 steps, and signal averaging = 4. The number of echoes used were 4000 (oxygenated Hb) and 2000 (oxidized Hb, deoxygenated Hb).



**Fig. 5** The  $T_1$ - $T_2$  correlational spectrum of blood microenvironment of (various hemoglobin variants) packed red blood cells. The variants were **(a)** HbE variant, **(b)** HbD variant, and **(c)** rare beta thalassemia variant, and other Hb variants (in Supplementary Fig. 3b). The zoom-in details of decomposed relaxation reservoirs for fast relaxation components ( $S$ -peak and  $T$ -peak) and the slow relaxation component (bulk water molecules,  $R$ -peak) is not shown. The coordinate for  $R$ -peak is indicated at upper left of the spectrum. The coordinate is represented by ( $T_2$  relaxation (in ms),  $T_1$  relaxation (in ms),  $A$ -ratio). The experimental parameters used were echo time = 200  $\mu$ s, number of echoes = 4000,  $T_1$ -incremental steps = 32 steps, and signal averaging = 4. Note that there is a possible artifact denoted as (\*).

HbE variant ( $T_2 = 0.94$  ms,  $T_1 = 9.44$  ms) and the beta thalassemia variant ( $T_2 = 0.56$  ms,  $T_1 = 10$  ms), in agreement with commonly observed clinical phenotype such as mild hemolysis due to increased oxidative damage. Interaction of Hb variants and other forms of hemoglobinopathies can lead to complex thalassemia syndromes with varying clinical phenotypes (Fig. 2).

**Machine learning assisted medical decision.** The 32 anonymized subjects consist of mixture of non-disease samples (wild-type), and disease samples (details in Supplementary Fig. 4). The NMR correlational spectroscopy maps ('molecular fingerprint') were converted into computer language for deep image analysis using statistical programming languages (e.g., R, Orange 3.1.2). Structural abnormalities in hemoglobin variants also lead to the observation of clinical methemoglobinemia in the late stage. The oxidized Hb samples were simulated examples for clinical methemoglobinemia.

The unsupervised learning techniques were used for dimension reduction (e.g., MDS), and classification (e.g., hierarchical

clustering) to assist in making medical decision (Fig. 6a). The 2D NMR correlational spectroscopy maps are complex 3D contour plots, and MDS technique was used to reduce higher dimension into two dimensional scatter plot which is more user-friendly for interpretation of information (Fig. 6b). Each feature ('molecular fingerprint' of one subject) was classified based on the common similarity within their intra-cluster as opposed to their inter clusters. Subjects were successfully classified into two clusters (disease (oxidized Hb, blue), non-disease (healthy wild type, red)) using the MDS technique ( $P < 0.05$ ), apart from the mutated counterpart (Hb variants, orange). In addition, the disease subtypes (sub-type 1: oxidized Hb, sub-type 2: partially oxidized Hb) were also observed (Fig. 6b). Distances between each subjects were shown in the heat map (Fig. 6c). Using hierarchical clustering, disease staging, prognosis or risk factor prediction (high/low-risk factor) were enabled (Fig. 6c, d). Other techniques (e.g., Isomap, linearly local embedding, t-sne) were evaluated and similar results were reproduced qualitatively (Supplementary Fig. 5).



**Blinded test: machine vs human learning.** The 32 anonymized subjects consist of mixture of non-disease samples (wild-type) and disease samples (details in Supplementary Fig. 4). Supervised learning models (e.g., logistic regression, neural network, k nearest neighbors (kNN) and naïve Bayes) were used to evaluate its' efficiency against human-being. K-fold cross validation (e.g.,  $k = 2, 3, 5$ ) and leave-one-out method were used for samplings. Five technicians were trained to differentiate between (diseases, non-disease) and subsequently were asked to classify the state of the spectrum based on a binary decision (diseases, non-disease) in blinded manner. At the end of the experiment, the results were cross-checked and classified as true positive, true negative, false positive and false negative (Supplementary Fig. 6). On-average, the machine learning models (e.g., CA = 0.885, sensitivity = 0.885, specificity = 0.887) outperformed the human being (e.g., CA = 0.794, sensitivity = 0.811, specificity = 0.720) in many aspects, when  $k = 5$  (Table 3). The performance of the supervised models, in general, improved with increasing value of  $k$  and achieved the maximum point when 'leave-one-out' method was used in training the datasets (details in Supplementary Fig. 7). Noticeably, the performance variation between each individual was larger than that of machine learning models as a result of human subjective judgment. On-average, machine learning

models (30 s) also took much shorter time than human (about 10 min) to complete the tasks given.

## Discussion

In this work, we showed that detailed and specific molecular microenvironment of water-proton interactions in blood can be mapped out using the two-dimensional  $T_1$ - $T_2$  correlational spectroscopy (Supplementary Table 1 and Supplementary Table 2). Interestingly, as water is ubiquitous to life form, water-protein interactions (e.g., the protein hydration) attracted considerable interests from terahertz spectroscopy<sup>44</sup> to neutron scattering<sup>45</sup>, provides an equivalent of 'inverse proteomic' information. This adds a new dimension to the existing traditional omics framework (e.g., genomic, proteomic) potentially revealing many biological pathways and understanding of fundamental of biological processes which have never been examined before.

It is demonstrated that the proposed technique here is capable of rapid label-free phenotyping the biological fluids in various physiological conditions (e.g., de/oxygenation level) and pathological states (e.g., blood oxidation, hemoglobinopathies) in uniquely personalized manner. We showed that time-to-result could be accomplished in minutes (Supplementary Fig. 1). With the recent availability of ultrafast signal acquisition methods<sup>12</sup> and efficient inversion algorithm<sup>7</sup>, real-time characterization and monitoring is possible. Aided with machine learning techniques, complicated NMR correlational maps were immediately transformed into clinically meaningful and user-friendly information.

Secondly, encoding multidimensional biochemical and biophysical information at molecular level using two-dimensional relaxation profiling (instead of chemical shifts), circumvent the limitation of using conventional big footprint NMR. Unlike high-field NMR spectroscopy, mass spectrometry, high-performance liquid chromatography where the instrumentation are often bulky and expensive (Table 4), an interesting NMR-based POCT proposed in this work offers inexpensive assay and instrumentation (e.g., open source code software-defined-radio<sup>46-48</sup>). Importantly, the unique and specific molecular fingerprint of liquid biopsy is able to provide a multiple global snapshot for disease dynamic monitoring in a minimally invasive manner<sup>49,50</sup>.

In summary, a novel concept of high unique and specific 'molecular fingerprint' of blood was demonstrated using time-domain two-dimensional NMR-based POCT. The assessment of multidimensional relaxation components of the blood was shown to be highly time- and patient-specific, delivering personalized information that is critical in clinical diagnostic, monitoring and prognostic purposes. Such personalized and precise method laid a strong foundation for the next generation of personalized

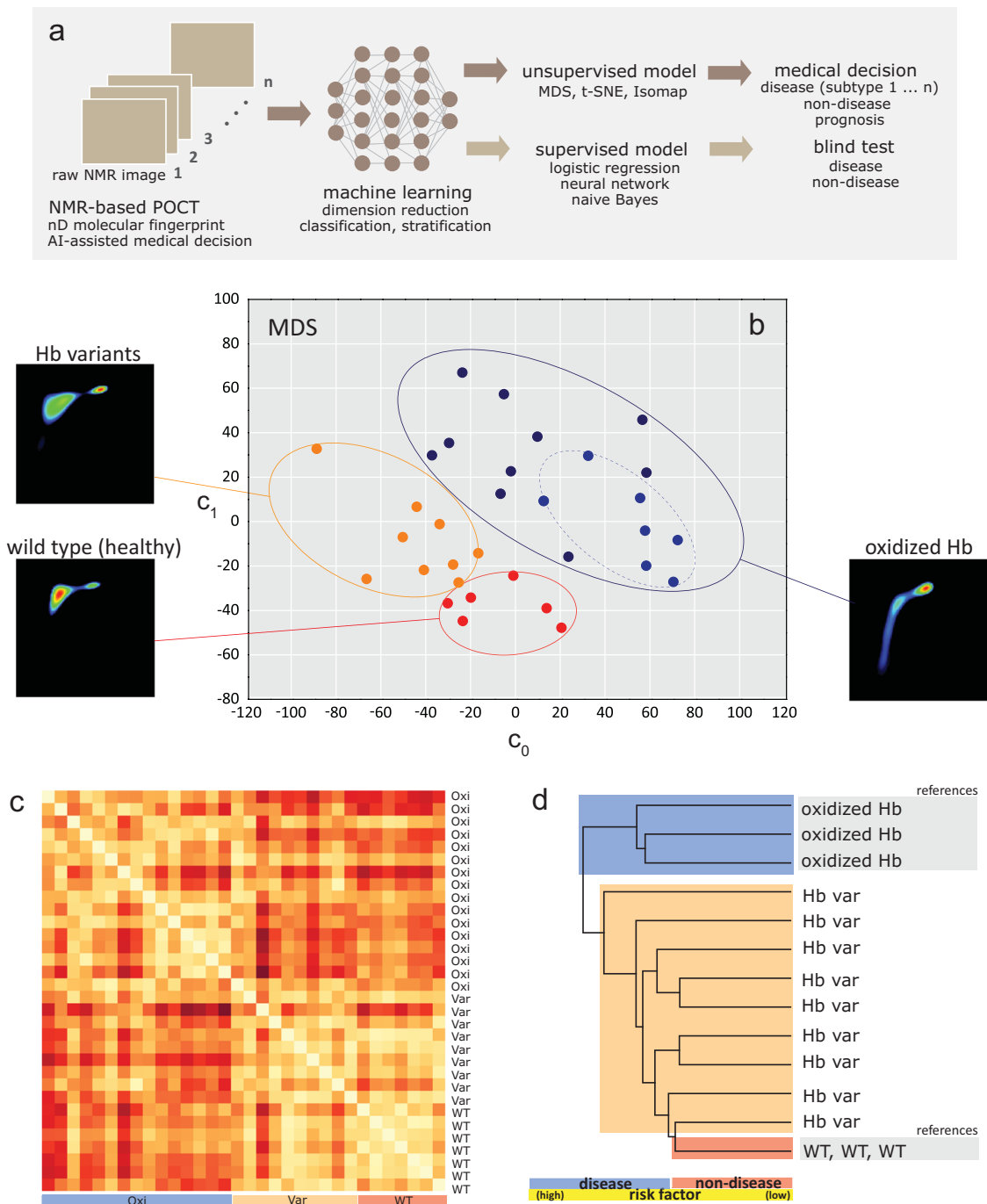
**Table 2 The decomposed relaxation reservoirs of packed red blood cells with the hemoglobin variants in (a) wild-type (control), (b) HbE variant, (c) HbD variant, and (d) rare beta thalassemia variant.**

Bulk water, R-peak	$T_1$ (ms)	$T_2$ (ms)	A ratio
a. Hb wild-type (control)	562	141	3.99
b. HbE variant	631	158	3.99
c. HbD variant	640	165	3.88
d. rare Hb variant	640	165	3.88
Hydration layer, S-peak	$T_1$ (ms)	$T_2$ (ms)	A-ratio
a. Hb wild-type (control)	335	4.47	74.94
b. HbE variant	335	4.22	79.39
c. HbD variant	373	4.18	89.23
d. rare Hb variant	362	6.48	55.86
Direct bound macromolecules, T-peak	$T_1$ (ms)	$T_2$ (ms)	A-ratio
a. Hb wild-type (control)	188	1.12	167.9
b. HbE variant	106	1.06	100
c. HbD variant	96.3	1.20	80
d. rare Hb variant	172	1.40	122.9

**Table 3 The performance of supervised machine learning models (e.g., neural network, k nearest neighbor (kNN), and logistic regression) in comparison to 5 technicians.**

Methods	AUC	CA	Sensitivity	Specificity	Precision	F1
Neural network	0.92	0.906	0.906	1	0.938	0.913
kNN	0.912	0.844	0.844	0.83	0.881	0.855
Logistic regression	0.927	0.906	0.906	0.83	0.914	0.909
Average	0.920	0.885	0.885	0.887	0.911	0.892
Technician 1	-	0.781	0.778	0.800	0.955	0.857
Technician 2	-	0.750	0.731	0.833	0.950	0.826
Technician 3	-	0.813	0.846	0.667	0.917	0.880
Technician 4	-	0.813	0.885	0.500	0.885	0.885
Technician 5	-	0.813	0.815	0.800	0.957	0.880
Average	-	0.794	0.811	0.720	0.932	0.866

The k-fold cross validation sampling methods (e.g.,  $k = 2, 3, 5$ ) and leave-one-out method were used to test and train the data. The performance of naïve Bayes model is well below the average human being (details in Supplementary Fig. 7). The abbreviations used were area under the curve (AUC), classification accuracy (CA), and F1-score is the harmonic mean for precision and sensitivity.



**Fig. 6 Machine-learning assisted NMR-based POCT in making medical decision.** **a** The workflow of machine learning in processing the complicated data into user-friendly medical decision (e.g., disease subtyping). The maps were converted into machine language using the image embedding (e.g., Squeeze Net) features. Dimensionality reductions were performed using various unsupervised models (e.g., MDS, t-SNE, Isomap). Supervised learning models (e.g., neural network, logistic regression, naive Bayes) were used to train and predict the data. The performance of supervised learning techniques were compared to that of human performance (Table 3). **b** The classification of three states (disease, non-disease, variants) and disease subtyping (sub-type 1: oxidized Hb, sub-type 2: partially oxidized Hb), and **c** heat map of 32 anonymized subjects processed using multidimensional scaling technique (300 max iterations, PCA-Torgersen). The legend (red, white) indicates (longer, shorter) distance between subjects. Other unsupervised models (e.g., linearly local embedding, Isomap, t-sne) were also evaluated for comparison (Supplementary Fig. 5). **d** The hierarchical clustering enabled disease staging, prognosis or risk factor prediction (high/low-risk subject) with respect to standard reference. For simplicity, three referencing states (WT and oxidized Hb) were shown. The non-disease state consists of (healthy wild-type), and disease state consist of (oxidized Hb, Hb variants). The short forms used were wild type (WT), oxidized Hb (Oxi), and Hb variants (Var). The clustering circles (dotted lines) were drawn for eye-balling purposes. The NMR correlational map of each subject is shown in Supplementary Fig. 4.

**Table 4 Comparison between the current proposed NMR-based POCT with AI-aided technology and existing state-of-the-art technologies (e.g., high-field NMR<sup>52</sup>, electrophoresis<sup>53</sup>, hplc<sup>54</sup>, and PCR-based assay<sup>45</sup>) which have been reported for clinical hemoglobinopathies.**

	NMR-based POCT	High-field NMR	Electrophoresis	HPLC cation exchange	PCR-based assay
Information	Phenomic	Proteomic	Proteomic	Proteomic	Genomic
Mode of action	Water-protein interactions (time domain)	Chemical shift (frequency domain)	Electric field (protein charge)	Mass transfer	DNA amplification
Multi-dimensional	Yes	Yes	Yes	No	No
AI-assisted	Yes	No	No	No	No
Equipment size	Bench-top	Large	Bench-top	Large	Large
Equipment price	Cheap	Expensive	Cheap	Expensive	Expensive
Price per assay	Ultra cheap	Medium	Medium	Expensive	Expensive
Sample processing	Easy	Easy	Difficult	Difficult	Difficult
Time to results	Minute	Hours	Hours	Hours	Hours
POCT	Yes	No	Yes	No	No
Non-destructive	Yes	Yes	No	No	No
Functional test	Yes	Yes	No	No	No
References	This work	Levitt et. al.,	Kutlar et. al.,	Kehra et. al.,	Kutlar et. al.,

The price per assay for NMR-based POCT refers to a single microcapillary tube (<\$0.10). In addition, the proposed method is label-free and therefore no chemical treatment is required. HPLC high-performance liquid chromatography.

medicine. The rapid, high-throughput and label-free nature of the proposed method has major implication in in vitro disease diagnosis and monitoring whereby the use of minimal invasive liquid biopsy read-out allows frequent testing. The use of machine learning algorithm improves the delivery of information (e.g., speed and accuracy), which may become a key factor in speeding up the translational of technological innovations to clinical routine and practices.

## Methods

**NMR setup and parameters.** The <sup>1</sup>H magnetic resonance measurements of bulk packed red blood cells were carried out at the resonance frequency of 21.57 MHz using a portable permanent magnet (Metrolab Instruments, Switzerland), B<sub>0</sub> = 0.5 T using a benchtop-type console (Kea Magritek, New Zealand). A temperature controller was set to maintain the measurement chamber at 24.5 °C. The T<sub>1</sub>-T<sub>2</sub> correlational pulse sequences were set at standard inversion recovery, followed by Carr–Purcell–Meiboom–Gill (CPMG) train pulses (Fig. 1).

The experimental parameters used; echo time = 200 μs, number of echoes = 2000 (for oxidized, and deoxygenated state) and 4000 (for oxygenated state), T<sub>1</sub> incremental steps = 32 (logarithmic) steps, and signal averaging = 4. A recycle delay of 2 s was set between each experiment to provide sufficiently long time to allow all the molecular spins to return to thermal equilibrium. The total acquisition time depends on the combination of a number of factors (e.g., number of scans, T<sub>1</sub>-incremental).

We demonstrated that a total experimental time in less than 6 min is sufficient for a high sensitivity and good spectral resolution, and without losing the spectral integrity (details in Supplementary Fig. 1). The 2D correlation maps were processed using built-in ILT algorithm (FISTA inversion)<sup>51</sup> method with 5000 iterations and smoothing parameter of 1 were used. The inversion typically completed in less than 2 min using a desktop computer (Intel Core Pentium i3 CPU @ 3.2 GHz, 1.74 Gb RAM).

**Clinical ethics and protocols.** This study received ethics approval from the local Institutional Review Board of the National Healthcare Group. K2 EDTA-anticoagulated whole blood samples were washed and re-suspended with phosphate buffer saline (PBS). Informed consent was obtained from all subjects involved in this study. All blood samples were either used immediately or kept at 4 °C and used within three to four days (unless mentioned otherwise) of collection before the micro MR analysis. To induce the Hb into various derivative states, the blood samples were incubated with the desired chemical as mentioned in the Text (e.g., sodium nitrite) and finally washed to remove the chemical residual. Heparinized micro capillary tubes (Fisher Scientific, PA) were used to transfer the processed blood and finally spun down at 6000 × g for 1 min to obtain packed red blood cells for MR measurements.

**Machine learning algorithm and workflow.** The NMR-based POCT can be used with or without the assistant of AI (Fig. 2). Machine learning techniques were used to transform the human complicated data (e.g., 2D NMR correlational maps) into user-friendly medical decision making following the workflow developed (Fig. 6). The maps

were converted into machine language using the image embedding features (e.g., Squeeze Net). Machine learning techniques were used to perform dimension reduction using various techniques (e.g., MDS, t-SNE, Isomap) (Supplementary Fig. 5).

**Blinded test.** Supervised learning models (e.g., neural network, k nearest neighbor, logistic regression, and naïve Bayes) were used to train and predict the data. We first trained 5 human beings to differentiate between (diseases, non-disease) and asked them to classify 32 anonymized subjects that were not seen before (Supplementary Fig. 4). They were allowed to backtrack (and change) the results as long as it was within the allocated time-frame (10 min). At the end of the experiment, the results were cross-checked and classified them as true positive (TP), true negative (TN), false positive (FP) and false negative (FN) (Supplementary Fig. 6). Statistical programming languages (e.g., Orange 3.1.2) was used for machine learning algorithm running on a personal laptop (Intel Core Pentium i7 CPU @ 2.70 GHz, 8.00 GB RAM). Once the models in machine learning were built, the run test takes less than 30 s to complete all the tasks, while each of the human beings took about 10 min on-average.

**Statistics and reproducibility.** Two tailed Student's *t* test was used to calculate the *P* value.

**Reporting summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this article.

## Data availability

The machine learning algorithms and 2D NMR raw maps along with any remaining info are available from corresponding author upon reasonable request at [weng.kung@inl.int](mailto:weng.kung@inl.int).

## Code availability

Machine learning calculations were made possible with Orange Data Mining. Full documentation, source code, and installation instructions are publicly available at <https://orange.biolab.si>.

Received: 15 July 2020; Accepted: 28 August 2020;

Published online: 28 September 2020

## References

- Salzmann, M., Pervushin, K., Wider, G., Senn, H. & Wuthrich, K. TROSY in triple-resonance experiments: new perspectives for sequential NMR assignment of large proteins. *Proc. Natl Acad. Sci. USA* **95**, 13585–13590 (1998).
- Bollard, M. E., Stanley, E. G., Lindon, J. C., Nicholson, J. K. & Holmes, E. NMR-based metabolomic approaches for evaluating physiological influences on biofluid composition. *NMR Biomed.* **18**, 143–162 (2005).
- Duarte, I. F., Diaz, S. O. & Gil, A. M. NMR metabolomics of human blood and urine in disease research. *J. Pharm. Biomed. Anal.* **93**, 17–26 (2014).



4. Viant, M. R., Lyeth, B. G., Miller, M. G. & Berman, R. F. An NMR metabolomic investigation of early metabolic disturbances following traumatic brain injury in a mammalian model. *NMR Biomed.* **18**, 507–516 (2005).
5. Ronconi, L. & Sadler, P. J. Applications of heteronuclear NMR spectroscopy in biological and medicinal inorganic chemistry. *Coord. Chem. Rev.* **252**, 2239–2277 (2008).
6. Hills, B. P. Applications of low-field NMR to food science. in *Annual reports on NMR spectroscopy* vol. 58 177–230 (Elsevier, 2006).
7. Song, Y.-Q. et al. T1–T2 correlation spectra obtained using a fast two-dimensional Laplace inversion. *J. Magn. Reson.* **154**, 261–268 (2002).
8. Haun, J. B. et al. Micro-NMR for rapid molecular analysis of human tumor samples. *Sci. Transl. Med.* **3**, 71ra16–71ra16 (2011).
9. Peng, W. K. et al. Micromagnetic resonance relaxometry for rapid label-free malaria diagnosis. *Nat. Med.* **20**, 1069–1073 (2014).
10. Peng, W. K., Chen, L. & Han, J. Development of miniaturized, portable magnetic resonance relaxometry system for point-of-care medical diagnosis. *Rev. Sci. Instrum.* **83**, 095115 (2012).
11. Veiga, M. I. & Peng, W. K. Rapid phenotyping towards personalized malaria medicine. *Malar. J.* **19**, 68 (2020).
12. Ahola, S. & Telkki, V.-V. Ultrafast two-dimensional NMR relaxometry for investigating molecular processes in real time. *ChemPhysChem* **15**, 1687–1692 (2014).
13. Anders, J., SanGiorgio, P. & Boero, G. An integrated CMOS receiver chip for NMR-applications. In *2009 IEEE Custom Integrated Circuits Conference* 471–474 <https://doi.org/10.1109/CICC.2009.5280786>. (IEEE, 2009).
14. Lee, H., Sun, E., Ham, D. & Weissleder, R. Chip–NMR biosensor for detection and molecular analysis of cells. *Nat. Med.* **14**, 869–874 (2008).
15. Dupré, A., Lei, K.-M., Mak, P.-I., Martins, R. P. & Peng, W. K. Micro- and nanofabrication NMR technologies for point-of-care medical applications – a review. *Microelectron. Eng.* **209**, 66–74 (2019).
16. Takeda, K. OPENCORE NMR: open-source core modules for implementing an integrated FPGA-based NMR spectrometer. *J. Magn. Reson.* **192**, 218–229 (2008).
17. Sun, N., Liu, Y., Lee, H., Weissleder, R. & Ham, D. CMOS RF biosensor utilizing nuclear magnetic resonance. *IEEE J. Solid-State Circuits* **44**, 1629–1643 (2009).
18. Ehrmann, K. et al. Microfabricated solenoids and Helmholtz coils for NMR spectroscopy of mammalian cells. *Lab Chip* **7**, 373 (2007).
19. Olson, D. L., Peck, T. L., Webb, A. G., Magin, R. L. & Sweedler, J. V. High-resolution Microcoil 1H-NMR for mass-limited, nanoliter-volume samples. *Science* **270**, 1967–1970 (1995).
20. Kong, T. F., Peng, W. K., Luong, T. D., Nguyen, N.-T. & Han, J. Adhesive-based liquid metal radio-frequency microcoil for magnetic resonance relaxometry measurement. *Lab Chip* <https://doi.org/10.1039/C1LC20853E> (2012).
21. Kong, T. F. et al. Enhancing malaria diagnosis through microfluidic cell enrichment and magnetic resonance relaxometry detection. *Sci. Rep.* **11425** <https://doi.org/10.1038/srep11425> (2015).
22. Castro, C. M. et al. Miniaturized nuclear magnetic resonance platform for detection and profiling of circulating tumor cells. *Lab Chip* **14**, 14–23 (2014).
23. Liang, M. et al. Magnetic barcode assay for genetic detection of pathogens. *Nat. Commun.* **4**, 55–65 (2013).
24. Chung, H. J., Castro, C. M., Im, H., Lee, H. & Weissleder, R. A magneto-DNA nanoparticle system for rapid detection and phenotyping of bacteria. *Nat. Nanotechnol.* **8**, 369–375 (2013).
25. Han, J. & Peng, W. K. Reply to ‘Considerations regarding the micromagnetic resonance relaxometry technique for rapid label-free malaria diagnosis’. *Nat. Med.* **21**, 1387–1389 (2015).
26. Peng, W. K., Chen, L., Boehm, B. O., Han, J. & Loh, T. P. Molecular phenotyping of oxidative stress in diabetes mellitus with point-of-care NMR system. *bioRxiv* 565325 <https://doi.org/10.1101/565325> (2019).
27. Peng, W. K., Han, J. & Loh, T. P. Micro magnetic resonance relaxometry. *U.S. Patent Application No. 15/136,887* (2016).
28. Peng, W. K., Ng, T.-T. & Loh, T. P. Machine learning assistive rapid, label-free molecular phenotyping of blood with two-dimensional NMR correlational spectroscopy. Preprint at <https://www.biorxiv.org/content/10.1101/2020.06.20.162974v1> (2020).
29. Peng, W. K. Clustering NMR: Machine learning assistive rapid (pseudo) two-dimensional relaxometry mapping. Preprint at <https://www.biorxiv.org/content/10.1101/2020.04.29.069195v1> (2020).
30. Song, Y.-Q., Ryu, S. & Sen, P. N. Determining multiple length scales in rocks. *Nature* **406**, 178–181 (2000).
31. Otting, G. NMR studies of water bound to biological molecules. *Prog. Nucl. Magn. Reson. Spectrosc.* **31**, 259–285 (1997).
32. Mathur-De Vré, R. The NMR studies of water in biological systems. *Prog. Biophysics Mol. Biol.* **35**, 103–134 (1980).
33. Dortch, R. D., Horch, R. A. & Does, M. D. Development, simulation, and validation of NMR relaxation-based exchange measurements. *J. Chem. Phys.* **131**, 164502 (2009).
34. McDonald, P. J., Korb, J.-P., Mitchell, J. & Monteilhet, L. Surface relaxation and chemical exchange in hydrating cement pastes: a two-dimensional NMR relaxation study. *Phys. Rev. E* **72**, 011409 (2005).
35. Lores Guevara, M. A., Naranjo, J. C. G. & Mirabal, C. A. C. MR relaxation studies of hemoglobin aggregation process in sickle cell disease: application for diagnostics and therapeutics. *Appl. Magn. Resonan.* <https://doi.org/10.1007/s00723-018-1104-0> (2018).
36. Thompson, B. C., Waterman, M. R. & Cottam, G. L. Evaluation of the water environments in deoxygenated sickle cells by longitudinal and transverse water proton relaxation rates. *Arch. Biochem. Biophysics* **166**, 193–200 (1975).
37. Aime, S., Fasano, M., Paoletti, S., Arnelli, A. & Ascenzi, P. NMR relaxometric investigation on human methemoglobin and fluoromethemoglobin. an improved quantitative in vitro assay of human methemoglobin. *Magn. Reson. Med.* **33**, 827–831 (1995).
38. Peng, W. K., Samoson, A. & Kitagawa, M. Simultaneous adiabatic spin-locking cross polarization in solid-state NMR of paramagnetic complexes. *Chem. Phys. Lett.* <https://doi.org/10.1016/j.cplett.2008.06.027> (2008).
39. Moallempour, M. et al. Methemoglobin effects on coagulation: a dose-response study with HBOC-200 (Oxyglobin) in a thrombelastogram model. *J. Cardiothorac. Vasc. Anesthesia* **23**, 41–47 (2009).
40. Bradley, W. G. MR appearance of hemorrhage in the brain. *Radiology* **189**, 15–26 (1993).
41. Lee, S. Y. et al. Evaluation of thalassaemia screening tests in the antenatal and non-antenatal populations in Singapore. *Ann. Acad. Med. Singap.* **48**, 5–15 (2019).
42. Yasmeen, H., Toma, S., Killeen, N., Hasnain, S. & Foroni, L. The molecular characterization of Beta globin gene in thalassaemia patients reveals rare and a novel mutations in Pakistani population. *Eur. J. Med. Genet.* **59**, 355–362 (2016).
43. Fucharoen, S., Siritanaratkul, N. & Wasi, P. Clinical manifestation of  $\alpha$ -Thalassaemia/Hemoglobin E Disease. *J. Pediatr. Hematol. Oncol.* **22**, 6 (2000).
44. Zhong, D., Pal, S. K. & Zewail, A. H. Biological water: a critique. *Chem. Phys. Lett.* **503**, 1–11 (2011).
45. Svergun, D. I. et al. Protein hydration in solution: experimental observation by x-ray and neutron scattering. *Proc. Natl Acad. Sci. USA* **95**, 2267–2272 (1998).
46. Peng, W. K. & Paesani, D. Omics meeting omics: towards the next generation of spectroscopic-based technologies in personalized medicine. *JPM* **9**, 39 (2019).
47. Asfour, A., Raouf, K. & Yonnet, J.-P. Software defined radio (SDR) and direct digital synthesizer (DDS) for NMR/MRI instruments at low-field. *Sensors* **13**, 16245–16262 (2013).
48. Hasselwander, C. J., Cao, Z. & Grissom, W. A. gr-MRI: A software package for magnetic resonance imaging using software defined radios. *J. Magn. Reson.* **270**, 47–55 (2016).
49. Yip, S. et al. A Canadian guideline on the use of next-generation sequencing in oncology. *Curr. Oncol.* **26**, e241–e254 (2019).
50. Alix-Panabières, C. & Pantel, K. Challenges in circulating tumour cell research. *Nat. Rev. Cancer* **14**, 623–631 (2014).
51. Zhou, X., Su, G., Wang, L., Nie, S. & Ge, X. The inversion of 2D NMR relaxometry data using L1 regularization. *J. Magn. Reson.* **275**, 46–54 (2017).
52. Arjmand, M. et al. Nuclear magnetic resonance-based screening of thalassaemia and quantification of some hematological parameters using chemometric methods. *Talanta* **81**, 1229–1236 (2010).
53. Kutlar, F. Diagnostic approach to hemoglobinopathies. *Hemoglobin* **31**, 243–250 (2007).
54. Khera, R., Singh, T., Khuana, N., Gupta, N. & Dubey, A. P. HPLC in characterization of hemoglobin profile in thalassaemia syndromes and hemoglobinopathies: a clinicohematological correlation. *Indian J. Hematol. Blood Transfus.* **31**, 110–115 (2015).

## Acknowledgements

This research was supported by the International Iberian Nanotechnology Laboratory (INL Start Up Grant (S4000040) and INL Seed Grant). T.T.N. acknowledges the support of A\*STAR AGA that provides access to a pool of very talented interns through its student attachment programme. W.K.P. would like to personally thank L. Daniel, an internship student along with T.T.N.'s co-supervision. L. Daniel conducted various preliminary MR measurements, and modifying the ILT code using MATLAB. We acknowledge Chin Hin Ng of Department of Oncology, National University of Hospital, Singapore for assistance in obtaining approval from the National Health Group (NHG) Institution Review Board.

## Author contributions

W.K.P. conceived the original idea, wrote the first draft of the paper, designed the experiments/protocols (e.g., NMR correlational experiments, machine learning coding), built the entire hardware setup and performed micro MR measurements. T.T.N. contribute in modifications on MATLAB code for ILT analysis. T.P.L. kindly provide the blood samples and engaged in various discussion on translational clinical aspects.

**Competing interests**

The authors declare no competing interests.

**Additional information**

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s42003-020-01262-z>.

**Correspondence** and requests for materials should be addressed to W.K.P. or T.P.L.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020