

# SCIENTIFIC REPORTS



OPEN

## Introducing Discrete Frequency Infrared Technology for High-Throughput Biofluid Screening

Caryn Hughes<sup>1,2</sup>, Graeme Clemens<sup>2</sup>, Benjamin Bird<sup>3</sup>, Timothy Dawson<sup>4</sup>, Katherine M. Ashton<sup>4</sup>, Michael D. Jenkinson<sup>5</sup>, Andrew Brodbelt<sup>5</sup>, Miles Weida<sup>3</sup>, Edeline Fotheringham<sup>3</sup>, Matthew Barre<sup>3</sup>, Jeremy Rowlette<sup>3</sup> & Matthew J. Baker<sup>2</sup>

Received: 25 June 2015  
Accepted: 23 December 2015  
Published: 04 February 2016

Accurate early diagnosis is critical to patient survival, management and quality of life. Biofluids are key to early diagnosis due to their ease of collection and intimate involvement in human function. Large-scale mid-IR imaging of dried fluid deposits offers a high-throughput molecular analysis paradigm for the biomedical laboratory. The exciting advent of tuneable quantum cascade lasers allows for the collection of discrete frequency infrared data enabling clinically relevant timescales. By scanning targeted frequencies spectral quality, reproducibility and diagnostic potential can be maintained while significantly reducing acquisition time and processing requirements, sampling 16 serum spots with 0.6, 5.1 and 15% relative standard deviation (RSD) for 199, 14 and 9 discrete frequencies respectively. We use this reproducible methodology to show proof of concept rapid diagnostics; 40 unique dried liquid biopsies from brain, breast, lung and skin cancer patients were classified in 2.4 cumulative seconds against 10 non-cancer controls with accuracies of up to 90%.

Key objectives for new diagnostic systems for diseases such as cancer include improving patient outcome through the identification of earlier stages, monitoring drug resistance and identifying high-risk populations for tumour progression. New diagnostic technological advances are not only required to be highly reliable but must also conform to cost-effective and high-throughput standards due to the drivers of ever-increasing workloads and rising costs in the medical profession.

Mid-infrared spectroscopy harnesses infrared radiation in the  $\nu$  25–120 THz frequency range ( $\lambda$  2.5–12  $\mu\text{m}$ ,  $1/\lambda$  4000–830  $\text{cm}^{-1}$ ) of the electromagnetic spectrum. The technique enables global bio-molecular fingerprints of complex organic materials to be recorded in a non-destructive and stain-free process<sup>1</sup>.

Attenuated total reflection (ATR-FTIR) spectroscopy is highlighted as an appropriate approach for point-of-care applications due to the minimal sample preparation and ability to rapidly acquire a single sample spectrum (in the order of seconds). While the method has been proven to provide cancer discrimination using blood serum<sup>2</sup>, it does not offer high-throughput utility as it requires a liquid biopsy to be spotted and dried on the ATR crystal surface before the spectrum is recorded. The drying process can take several minutes and the crystal must be subsequently cleaned before a new sample is presented; as a consequence, storage of samples for repeat measurements is also not an option<sup>3</sup>.

Mid-infrared Fourier transform micro-spectroscopic (FTIRM) imaging has been suitably applied to complement histological, cytological and pathological studies to detect subtle changes in the proteome and metabolome of diseased cells and tissue. FTIRM is fast proving to be a reliable research tool with clinical utility for disease-state diagnostics<sup>4–8</sup>. The application of IR imaging for dried liquid blood biopsy serum samples, however, is a relatively new application<sup>3,9</sup>. By using an infrared imaging approach, a high-throughput screening application can be realised for the biomedical laboratory; liquid biopsies that are dispensed and dried to create multi-patient micro-arrays to deliver rapid data collection, quality control and classification processes<sup>3</sup>.

<sup>1</sup>University of Manchester, School of Chemical Engineering and Analytical Science, Manchester, M13 9PL, United Kingdom. <sup>2</sup>WestCHEM, Technology and Innovation Centre, 99 George St, Department of Pure and Applied Chemistry, University of Strathclyde, Glasgow, G1 1RD, UK. <sup>3</sup>Daylight Solutions Inc., San Diego, CA 92128, USA. <sup>4</sup>Lancashire Teaching Hospitals NHS Trust, Royal Preston Hospital, Department of Pathology, Preston, PR2 9HT, UK. <sup>5</sup>The Walton Centre NHS Foundation Trust, The Walton Centre for Neurology and Neurosurgery, Liverpool, L9 7LJ, UK. Correspondence and requests for materials should be addressed to C.H. (email: caryn.hughes@manchester.ac.uk @Dr\_Caryn\_Hughes) or M.J.B. (email: matthew.baker@strath.ac.uk, @ChemistryBaker)

Biomedical applications of FTIRM have been proved in research laboratories but despite a surge in technological advance and theoretical understanding<sup>10</sup>, FTIRM has yet to be fully integrated into large-scale clinical application. One of the technological caveats of the imaging technique is the requirement to scan the entire IR spectrum; a redundant approach as clinically relevant diagnostic markers are found at very specific frequencies. The infrared source on a benchtop FTIRM system is also low-throughput and can require longer scanning times to improve signal to noise. As a consequence, FTIRM in its current state of development is not considered to be an efficiently rapid and high-throughput technique per se.

The recent integration of high-brightness, broadly tuneable mid-infrared quantum cascade lasers (QCLs) into a microscope system may address these concerns. The combination of the QCL source, refractive-based high numerical aperture objectives and a large format detector system has enabled the option of high-definition diffraction-limited resolution, without a trade-off in signal to noise and field of view when compared to FTIR-based microscope systems. The technique is called discrete frequency infrared (DFIR) imaging as the laser is tuneable to a custom range, providing a substantial reduction in data acquisition time, creating potential for efficient, real-time data collection for disease-specific diagnostics on a clinically appropriate time scale<sup>11,12</sup>.

Results presented in this communication illustrate the developmental application and validation of DFIR with a focus on cancer detection using human blood serum. Initially, the use of an automated piezo-jetted serum dispenser and subsequent DFIR imaging of dried serum spots from human pooled serum are assessed to explore sample uniformity. The samples were imaged using different discrete frequency ranges and subsequently assessed after implementing data processing and quality control mechanisms. The advantage of rapid infrared imaging as opposed to point spectra is the ability to evaluate and monitor sample quality control requirements. Secondly, it was necessary to investigate whether the diagnostic capability of cancer detection was preserved when moving from full frequency (as would be collected using an FTIR system) to discrete frequency mode (resulting in a reduction in spectral information). Finally, the high-throughput diagnostic capability of DFIR technology is presented by a proof of concept example involving a series of samples from multiple patients presenting with different cancers.

## Results

**Investigating collection parameters and sample reproducibility.** Several studies have proven manual spotting of serum samples to be an adequate method of preparation for disease-state IR diagnostics<sup>2,13,14</sup>. In a clinical setting, however, an automated dispenser would need to be utilised to increase consistency in sample placement, volume accuracy and high-throughput sample processing as recent investigations have shown<sup>15</sup>. A total of 40 technical replicate serum spots from human pooled serum were used to investigate the feasibility of automated sample preparation and also to determine the consequence of DFIR parameter selection on spectral reproducibility.

The serum spots displayed fairly consistent morphological drying patterns, as illustrated by their 3D mesh plots displaying total intensity of infrared absorbance (Fig. 1). The samples were chemically imaged in a  $2 \times 2$  mosaic map using three different DFIR ranges at either 199, 14 or 9 specific frequencies (Table 1). Each replicate image was subject to a process of quality control by spatially locating pixel spectra of poor quality, which resulted from issues relating to sample thickness, signal to noise and scattering. For each replicate, the remaining spectra were averaged, creating a single high-quality spectrum per sample.

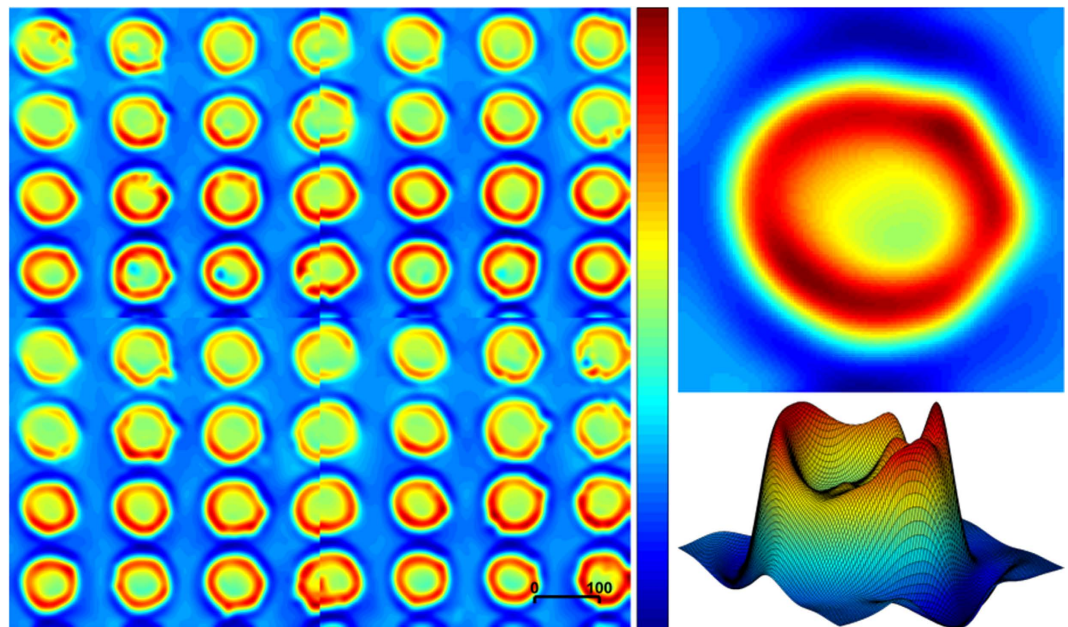
The mean variability for each dataset of the 40 replicate observations is shown in Table 1 and Fig. 2. The impact of the quality control regime and frequency range on spectral reproducibility indicated that the use of 14 discrete frequencies, chosen from the fingerprint range, provided the best compromise between speed of acquisition and reproducibility.

**Moving from full to sparse frequency: Diagnostic potential is maintained.** Non-cancerous serum samples have previously been discriminated from patients with brain tumours of increasing severities in low-grade to high grade gliomas with the use of ATR-FTIR spectroscopy and chemometric analytical techniques; Cancerous and non-cancerous blood serum could be classified with sensitivities and specificities of 94 and 97% respectively<sup>2</sup>. The discriminatory spectral features between the non-cancer and cancerous samples were found at frequencies largely associated with the amide I and II bands which correspond to C=O stretching, C-N stretching and N-H bending vibrational modes<sup>9</sup>.

In order to investigate whether the same discriminatory features could be found in the serum using rapid DFIR technology, serum samples from brain were analysed by using a quantum cascade laser (QCL) IR system for both full-frequency and discrete-frequency (DFIR) modes (specifically 14 frequencies in this case). Figure 3a,b display the mean representative spectrum recorded from serum samples of non-cancerous (green) and cancerous (blue) patients using full and sparse frequency data collection modes respectively. Despite the substantial reduction in data, the discrete frequency spectra exhibited recognisable bandshapes of their equivalent full frequency spectra.

Centroid position values were calculated for the amide I and II peaks and subsequently displayed by the ratio in the correlation plots in Fig. 3c,d for full and sparse frequency modes respectively, where differences were revealed; For the discrete frequency data, the average peak centroid values had shifted relative to the full frequency positions ( $1651$  to  $1645\text{ cm}^{-1}$  and  $1548$  to  $1552\text{ cm}^{-1}$  for the amide I and II peaks respectively). Inevitably, there was greater sensitivity in the intra-class distinction using the full data range mode, as data points belonging to individual patient sample spectra can be more clearly resolved. This is due to the higher accuracy in peak centroid calculation as a result of the additional data points contributing to peak band shape.

It was noted, however, that both full and discrete frequency modes resulted in incontestable discrimination of cancer versus non-cancer data overall, demonstrating that streamlined data acquisition via discrete frequency mode has an equally acceptable diagnostic power in the goal of rapid cancer detection. There are currently no



**Figure 1. Technical replicability of automated liquid biopsy dispensing.** Automated sample preparation of 56 dried biopsies from pooled human blood serum; A  $2 \times 2$  mosaic DFIR image displaying total intensity of infrared absorbance which indicates sample thickness (with increasing values from blue to red as shown in the colour bar). The automated dispenser allowed for the creation of smaller deposits relative to hand-dispensed samples; the scale bar displays 100 pixels (equivalent to a size of  $130 \mu\text{m}$ ). An individual sample is also displayed in a 2D and topographical mesh diagram showing uneven sample thickness, confirming the 'coffee-ring' drying phenomenon. The sample was imaged three times using 199, 14 and 9 discrete frequencies.

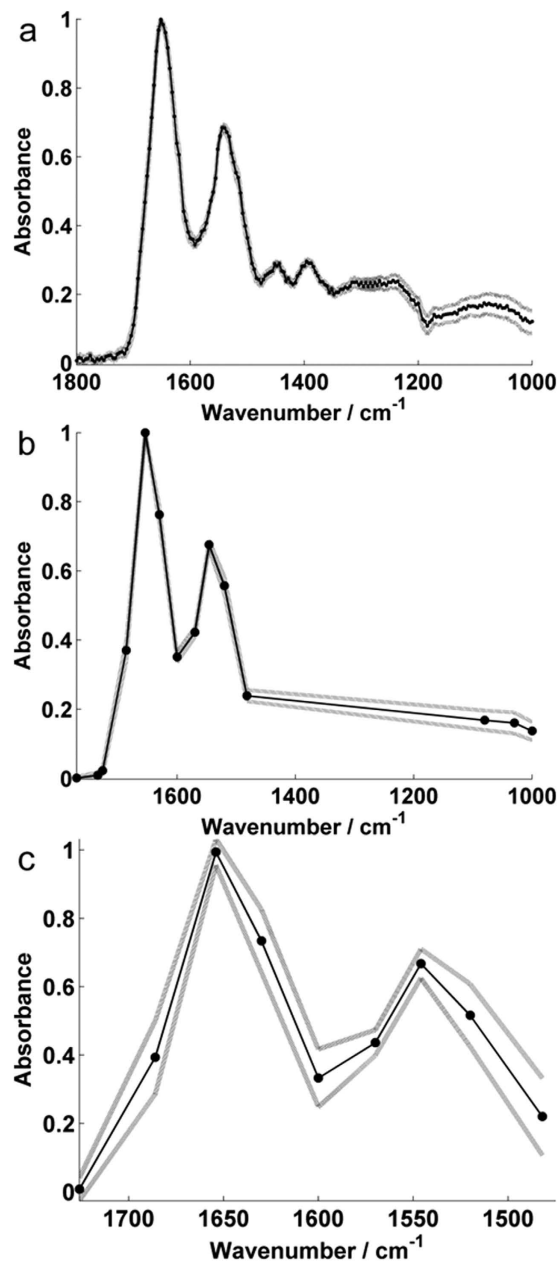
| Number of Discrete Frequencies | Discrete Frequencies*  | Acquisition Time** (min) | $\Delta\text{Abs}^+$ | RSD <sup>++</sup> (%) |
|--------------------------------|--|--------------------------|----------------------|-----------------------|
| 199                            | 1000–1800  | 45.2                     | 0.017                | 0.6                   |
| 14                             | 1000, 1030, 1080, 1482, 1520, 1546, 1570, 1600, 1630, 1654, 1686, 1726, 1734, 1770 | 6.8                      | 0.017                | 5.1                   |
| 9                              | 1482, 1520, 1546, 1570, 1600, 1630, 1654, 1686, 1726                               | 4.0                      | 0.071                | 15.0                  |

**Table 1. Data collection speeds and spectral reproducibility for the mosaic image (Figure 1) acquired using three different DFIR ranges.** \*Quoted in wavenumbers ( $\text{cm}^{-1}$ ). The  $1000\text{--}1800 \text{ cm}^{-1}$  range included data point spacing of  $4 \text{ cm}^{-1}$ . \*\*The data collection time for the ( $2 \times 2$  tile) image. +The average change in absorbance value across the discrete frequencies for the spectral dataset ( $n = 40$ ). ++The mean relative standard deviation for the spectral dataset ( $n = 40$ ).

techniques in the clinic for the discrimination of cancer vs non-cancer for multiple disorders. To compare and benchmark our results a study on breast mammography using 2 radiologists (240 women with breast cancer and 240 women with normal breast) reported sensitivities and specificities of 75.8 and 85.4% for radiologist 1 and 75.4 and 89.2% for radiologist 2 respectively<sup>16</sup>. This demonstrates that our results are of clinical acceptance, can be provided before and help direct an invasive procedure, but are also a direct physical measurement of the patient blood biochemistry. Conventional methods are reliant upon human interpretation of tissue biopsy morphology and thus often provide more sporadic levels of inter-observer accuracy within the clinic and can be inadequate to flag the early signs of brain disease. It is quite often the case that by the time symptomatic changes are addressed via imaging and biopsy, the disease is at an advanced stage and prognosis is poor. Thus, a serum based screening via infrared methods may in the future help elucidate earlier forms of disease and allow more successful intervention and treatment.

**Demonstrating proof of concept for DFIR rapid cancer detection.** Using the results of the preliminary investigation and knowledge from previous studies, an unprecedented proof of concept experiment was conducted to test the premise that DFIR imaging technology could rapidly and consistently lead to the correct classification of dried fluid-biopsies from multiple individuals that presented with different cancers.

A sample matrix of 50 human blood serum liquid biopsies were prepared by spotting onto a single slide and subsequently air-dried. The sample matrix was comprised of 10 individual observations from non-cancer



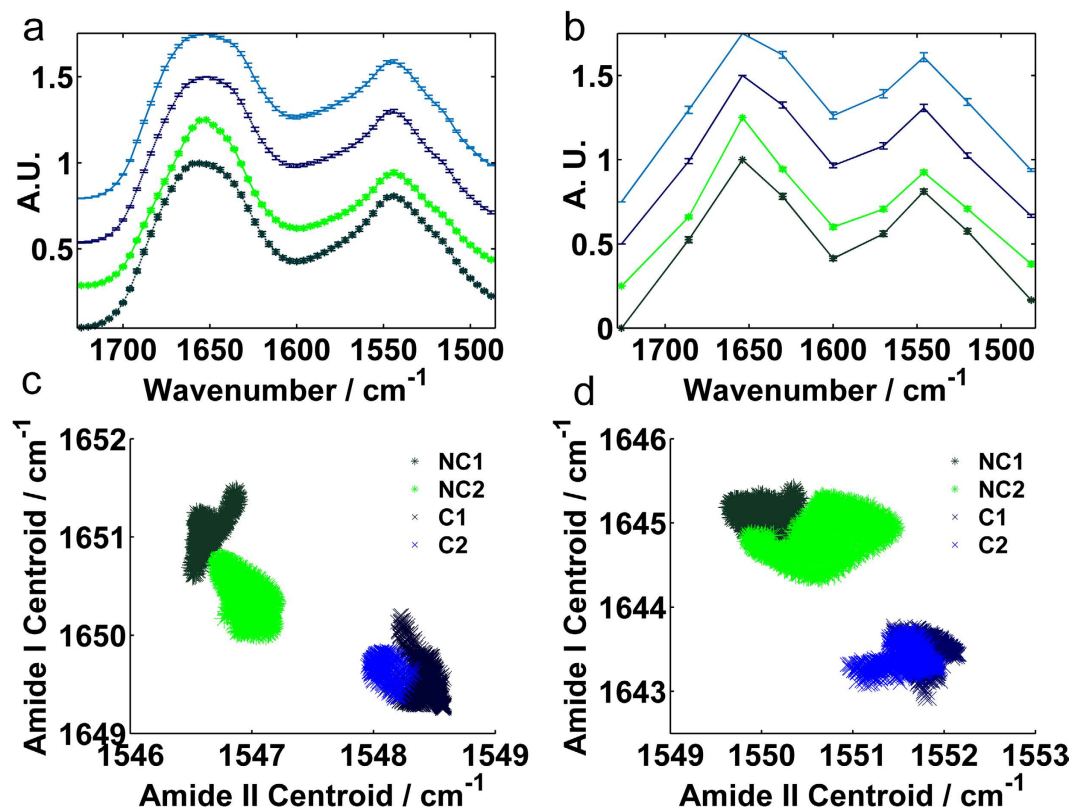
**Figure 2. Data processing replicability for different data collection frequencies.** Samples shown in Fig. 1 were imaged using different frequency ranges. A total of 40 of the technical replicate deposits were subject to individual data processing for each frequency range. Mean representative spectra from the 40 technical replicates  $\pm$  standard deviation are shown for (a) 199 (b) 14 and (c) 9 discrete frequencies.

controls, and brain, breast, lung and skin cancer. The sample was imaged as one DFIR image mosaic. For each patient, a single representative spectrum comprising of just 14 discrete frequencies was obtained.

The spectra were then tested for cancer detection using a radial-basis-function support vector machine (SVM) model<sup>3,14</sup>. Each cancer spectrum in its class was tested against a model built on a population of remaining cancer patient spectra and all control spectra. Each of the 40 cancer patient spectra were tested in triplicate, with slight variations in model parameters in the different iterations (SI 1). Each time the SVM model was trained, a single patient spectrum took 0.02 seconds to test. In almost all cases, the classification model was robust; over the 3 replicate testing runs, a unanimous decision was reached 95% of the time. For the two instances which resulted in different decision outcomes when tested in the triplicate, the modal outcome was taken (SI 1).

The concluding ‘cancer versus non cancer’ classification test had a 9/10 success rate for brain, breast and lung patients and an 8/10 rate for skin cancer patients (Fig. 4). It took a cumulative time of 2.4 seconds to test (in triplicate) the 40 cancer-patient samples.





**Figure 3.** Comparing diagnostic potential from full range to sparse frequency data. (a,b) mean representative spectra for non-cancer (NC, green, \*) and brain cancer (C, blue) patient samples. Error bars denote standard deviation (mean values of standard deviation for NC1, NC2, C1, C2 were 0.01, 0.009, 0.009, 0.008 for full and 0.011, 0.010, 0.012, 0.016 for sparse range respectively). (c,d) Corresponding Amide I and II peak centroid correlation plots for full and sparse frequency data collection modes respectively.

## Discussion

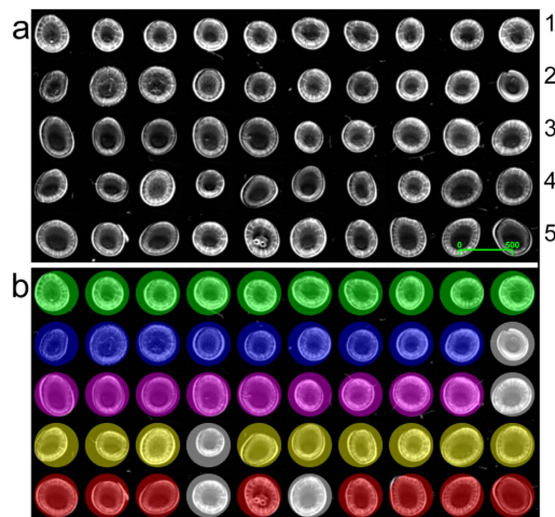
As with any new approach, it is important to consider parameter selection in order to achieve quality data, collected with maximum efficiency. One can see from Fig. 1 that even serum droplets created with an automated dispenser can exhibit the so-called ‘coffee-ring effect’<sup>3</sup>. The drying pattern of the serum can cause variation in sub-regional sample thickness and molecular composition, as well as causing physical light scattering perturbations that can contaminate the spectral signature of the sample<sup>3</sup>. Any difference in molecular composition across the sample can be managed by averaging the spectral data collected in each pixel associated with an individual serum spot. The spectral perturbations caused by influence of light scattering from a bio-fluid sample are far from acute, in comparison to biological cells for instance. Nevertheless, it is important to remove such affected spectra and spectra dominated by saturation due to non-linear absorption from the overall sample measurement.

The initial experiment compared the findings taken with both discrete frequency and full frequency (‘FTIR-like’) approaches from a small sample batch consisting of hundreds of pixel spectra. Despite the fact that a quality control test was not adopted at this stage (spectra were randomly selected from the centre of the serum spot images), both spectral datasets were in agreement; the scatter pattern of amide I/II centroid ratios convincingly distinguished cancer from non-cancer control data points. The pattern indicated that subtle changes in the proteome of the complex serum bulk (on a macroscopic scale) would be detected whether full or discrete frequency data collection mode was chosen.

A proof of concept study was performed to illustrate DFIR-led rapid diagnosis of individual fluid biopsies in multiple cancers. A mosaic DFIR hyperspectral image was constructed with 50 unique serum spot samples. The spectral outputs for the mosaic were systematically assessed under a quality control strategy before the representative spectrum of the individual serum spot was produced. Each cancer patient sample was blind-tested against a model built from respective cancer patient class data alongside control samples.

The testing process for a patient spectrum was performed with almost instantaneous speed at 0.02 seconds. It is true that the 93% reduction of number of required data points is certainly a factor in the classification speed (i.e. 14 data points compared to a full frequency spectrum equivalent (of the fingerprint range) with 199 data points). This is not, however, a result to champion in isolation given that a full frequency single spectrum would still be arguably very fast to classify.

One major advantage of acquiring DFIR data lies in the lighter computational requirements for pre-processing prior to the creation of single spectra. A simple method to improve big data management is to reduce the initial data set size. The mosaic image shown in Fig. 4 with 50 spots has dimensions of 2400 × 4800 pixels; the



**Figure 4. Rapid classification of multiple cancers using dried human serum liquid biopsies.** (a) Grayscale image mosaic of 50 unique serum biopsies (row 1 non-cancer controls, 2 brain 3 breast, 4 lung and 5 skin cancer patient samples) the scale bar displays 500 pixels (equivalent to a size of 2.1 mm). (b) Classification outcomes for correct cancer diagnosis were accurate to 90%, 90%, 90% and 80%, highlighted by colour for brain (blue), breast (pink), lung (yellow) and skin (red) cancer samples respectively.

unprocessed hyperspectral image will contain  $1.344 \times 10^7$  spectra. With only 14 discrete frequencies in DFIR mode, the total number of data points sampled would equate to  $1.882 \times 10^8$  (as a point of measure, the 14-DFIR mosaic image loaded into Matlab with  $1.882 \times 10^8$  data points required 1.83 GB of memory). If the image was created using a wider frequency range, such as 199 frequencies, this would increase to  $2.675 \times 10^9$  data points. Increasing the number of frequencies sampled therefore has a considerable effect on computational power and inevitable implications on speed during the spectral pre-processing and quality control steps.

The second advantage is the actual time taken to acquire the data. Using the collection parameters as described in the methods section, a single image of a single serum spot would take 11.3 minutes for  $1000\text{--}1800\text{ cm}^{-1}$  (199 data points), 1.7 minutes for 14 data points and 1 minute for 9 data points. (NB: There have already been recent advances made in the speed of data collection since these experiments were conducted; the laser can now be programmed to execute a sweep of wavelengths, sending triggers to the software to acquire frames when it has reached each wavelength, which is a much more efficient method to acquire the data and has subsequently reduced the time taken for the 199 data points to 4.3 minutes, for 14 points to 56 seconds and 9 points to 36 seconds).

Upon consideration of these results, further investigation of technical parameters and optimisation in sample preparation was considered. The use of an automated piezo dispensing system allowed for more consistency and produced smaller spots with a more consistent size due to spotting accuracy and volume control. As a consequence, this would also increase the high-throughput spot capacity as a result of the smaller sample spot area.

Considerable reduction of spectral acquisition time was possible by using 14 discrete frequencies (Table 1). The spots all originated from the same patient sample. In terms of intra-sample variability, despite the 4.5% increase in relative standard deviation, the average change in the overall spectral absorbance remained consistent with the full frequency range. The largest region of spectral variance resided in the lower frequency region (towards  $1000\text{ cm}^{-1}$ ). For spectral baseline purposes, it was useful to collect one or two data points at these lower frequencies. Once spectra have been scaled, however, variability can be reduced further by subsequent spectral truncation; especially relevant if the most salient features for the diagnostic test were found in the higher frequency (amide I and II) regions.

To meet high-throughput requirements of a clinical diagnostic tool, we have shown that automated sample preparation can increase technical replicate consistency and conformity, maximising the number of samples probed in the sampling area. In terms of data processing and accessibility, spectral apps are also currently in development that would allow mobile devices to access and manipulate spectra for real-time reporting<sup>17</sup>.

Given the full frequency studies have been thoroughly conducted, transition to DFIR is an obvious step towards building a rapid disease detection tool. By tuning the QCL to acquire data from frequencies of interest, DFIR imaging allows the opportunity to significantly increase the speed of data acquisition and processing. The indicative power of the three studies combined demonstrate the solid proof of concept that DFIR imaging is a plausible technological candidate to achieve a rapid, high-throughput tool for cancer diagnostics in the near future.

## Methods

**Ethical Approval.** Human blood serum patient samples were obtained from the Walton Research Tissue Bank (WRTB) and Brain Tumour North West (BTNW) Tissue Bank with informed consent from all subjects. Joint application for clinical samples for experiments performed under guidelines and regulations from the Brain

Tumour North West and the Walton Research Tissue Banks were approved by the BTNW/WRTB Committee (BTNW/WRTB 13\_01, Application Number 13\_01). Samples obtained represent high-grade gliomas (primary brain tumours) (WHO Grade III and IV) and skin, breast and lung tumours that have metastasised to the brain.

**Sample preparation.** Pooled human serum was purchased from TCS Biosciences, UK (0.2  $\mu\text{m}$  sterile filtered CS100–100). 56 technical replicate spots were dispensed using a piezo-driven auto-dispenser every 6.0 ms at the speed of 100 mm/sec and were left to dry in ambient conditions. The human serum of tissue bank origin was manually pipetted by hand in 0.5  $\mu\text{L}$  droplets and left to dry in ambient conditions.

**Data collection.** All spectral data was acquired using a QCL Spero™ microscope (Daylight Solutions Inc., San Diego, CA, USA). All data supporting this research are openly available from MJB. The authors direct the readers to Hughes *et al.*<sup>18</sup> for a published schematic of the instrument. Data was acquired in different ranges; either at 199, 14 or 9 frequencies (specific frequencies can be found in Table 1). The specific frequencies were chosen based upon the study by Hands *et al.*<sup>2</sup>. All data collected during this study was recorded using what we term, “Step and Measure Mode”, which is software driven. Briefly, the laser is commanded to step to a wavelength and an intensity image is acquired. Multiple frames can be acquired to increase signal-to-noise for low light level situations. The software then commands the laser to tune to the next wavelength in an iterative process.

The piezo-jetted drops were acquired using a  $2 \times 2$  mosaic with a high magnification objective (Fig. 1) and were ca. 100  $\mu\text{m}$  in diameter. (FOV = 650  $\mu\text{m} \times 650 \mu\text{m}$ , 12.5  $\times$  magnification, NA = 0.70, pixel size = 1.36, spatial resolution = ca. 5  $\mu\text{m}$  at  $\lambda$  5.5  $\mu\text{m}$ ). The hand dropped serum spots were ca. 1000  $\mu\text{m}$  in diameter acquired using single frames with a low magnification objective (FOV = 2 mm  $\times$  2 mm, 4  $\times$  magnification, numerical aperture (NA) = 0.15, pixel size = 4.25, spatial resolution = ca. 25  $\mu\text{m}$  at  $\lambda$  5.5  $\mu\text{m}$ ). The single frames created  $5 \times 10$  mosaic matrix shown in Fig. 3.

**Data processing and analysis.** For the initial experiment, serum spots were considered on an individual basis. The 56 serum spots were captured in a  $2 \times 2$ -frame mosaic IR image. A total of 40/56 spots were analysed for reproducibility, as some of the spots were not completely captured, or fell on the boundaries when the single tile images were stitched together (Fig. 1). Mean spot spectra were generated after the quality control process as described for the second experiment. Reproducibility was assessed by determining both the average change in absorbance and average relative standard deviation across each frequency range for the 40 spectral observations (Table 1, Fig. 2).

In order to compare the full and discrete frequency data, spectra were sampled from the centre of two brain cancer patient and two control serum spots that were manually spotted. Data was collected with both full and discrete frequency ranges (199 or 14 data points) and was subject to the same processing treatment by min-max scaling. The two datasets were compared by determining the peak centroid spectral moment for the amide bands in each individual spectrum, where differences in band frequency position and band shape will determine the centroid value. The centroid positions of the amide I and amide II band were correlated, illustrated in Fig. 1(c,d).

In the final experiment, a multi-mosaic image was constructed that comprised of 50 manually prepared serum spots. The spots represented five different classes; non-cancer controls versus, brain, breast, lung and skin cancer; all sample spots were sourced from 50 independent patient subjects. Image segmentation was applied using *k*-means cluster analysis to identify good quality spectra from each pixel that was not representative of either background (no or low signal) or distorted spectra (saturated data due to sample thickness) (SI 2). Quality-passed spectra were averaged to create a single representative spectrum per patient sample. A radial-basis function support vector machine (RBF-SVM) Model was used to demonstrate non-cancer/cancer proof of concept classification. Classification was done on a separate basis for each cancer class versus the non-cancer control set. The model was trained with the 10 control spectrum plus 9 of the cancer class spectra on a ‘leave-one-out’ basis, i.e. to test brain cancer patient 1 data, the model was training using patients 2–10. The training included 3-fold cross-validation to determine the SVM parameters (C and  $\gamma$ ) to gain the best training classification accuracy. The parameter values obtained for patient 1 were then fixed when repeating the training model to test patients 2–10 respectively, which resulted in either positive or negative classification. For robustness, each overall cancer model was trained in triplicate as the SVM parameters could be subject to value changes (SI 1).

## References

- Baker, M. J. *et al.* Using Fourier transform IR spectroscopy to analyze biological materials. *Nat. Protocols* **9**, 1771–1791, doi: 10.1038/nprot.2014.110.
- Hands, J. R. *et al.* Attenuated Total Reflection Fourier Transform Infrared (ATR-FTIR) spectral discrimination of brain tumour severity from serum samples. *Journal of Biophotonics* **7**, 189–199, doi: 10.1002/jbio.201300149 (2014).
- Hughes, C. *et al.* Assessing the challenges of Fourier transform infrared spectroscopic analysis of blood serum. *Journal of Biophotonics* **7**, 180–188, doi: 10.1002/jbio.201300167 (2014).
- Bhargava, R. Towards a practical Fourier transform infrared chemical imaging protocol for cancer histopathology. *Analytical and Bioanalytical Chemistry* **389**, 1155–1169, doi: 10.1007/s00216-007-1511-9 (2007).
- Bird, B. *et al.* Infrared spectral histopathology (SHP): A novel diagnostic tool for the accurate classification of lung cancer. *Laboratory Investigation* **92**, 1358–1373, doi: 10.1038/labinvest.2012.101 (2012).
- Hughes, C. *et al.* FTIR microspectroscopy of selected rare diverse sub-variants of carcinoma of the urinary bladder. *Journal of Biophotonics* **6**, 73–87, doi: 10.1002/jbio.201200126 (2013).
- Patel, I. I. *et al.* Infrared microspectroscopy identifies biomolecular changes associated with chronic oxidative stress in mammary epithelium and stroma of breast tissues from healthy young women: Implications for latent stages of breast carcinogenesis. *Cancer Biology and Therapy* **15**, 225–235, doi: 10.4161/cbt.26748 (2014).
- Walsh, M. J., Holton, S. E., Kajdacsy-Balla, A. & Bhargava, R. Attenuated total reflectance Fourier-transform infrared spectroscopic imaging for breast histopathology. *Vibrational Spectroscopy* **60**, 23–28, doi: 10.1016/j.vibspec.2012.01.010 (2012).

9. Clemens, G., Bird, B., Weida, M., Rowlette, J. & Baker, M. J. Quantum cascade laser-based mid-infrared spectrochemical imaging of tissues and biofluids. *SpectroscopyEurope* **26**, 14–19 (2014).
10. Bassan, P. *et al.* FTIR microscopy of biological cells and tissue: Data analysis using resonant Mie scattering (RMieS) EMSC algorithm. *Analyst* **137**, 1370–1377, doi: 10.1039/c2an16088a (2012).
11. Bassan, P., Weida, M. J., Rowlette, J. & Gardner, P. Large scale infrared imaging of tissue micro arrays (TMAs) using a tunable Quantum Cascade Laser (QCL) based microscope. *Analyst* **139**, 3856–3859, doi: 10.1039/c4an00638k (2014).
12. Yeh, K., Kenkel, S., Liu, J.-N. & Bhargava, R. Fast Infrared Chemical Imaging with a Quantum Cascade Laser. *Analytical Chemistry* **87**, 485–493, doi: 10.1021/ac5027513 (2015).
13. Gajjar, K. *et al.* Fourier-transform infrared spectroscopy coupled with a classification machine for the analysis of blood plasma or serum: A novel diagnostic approach for ovarian cancer. *Analyst* **138**, 3917–3926, doi: 10.1039/c3an36654e (2013).
14. Hands, J. R. *et al.* Investigating the rapid diagnosis of gliomas from serum samples using infrared spectroscopy and cytokine and angiogenesis factors. *Analytical and Bioanalytical Chemistry* **405**, 7347–7355, doi: 10.1007/s00216-013-7163-z (2013).
15. Ollesch, J. *et al.* FTIR spectroscopy of biofluids revisited: An automated approach to spectral biomarker identification. *Analyst* **138**, 4092–4102, doi: 10.1039/c3an00337j (2013).
16. Houssami, N. *et al.* The influence of clinical information on the accuracy of diagnostic mammography. *Breast Cancer Res Treat* **85**, 223–228, doi: 10.1023/b:brea.0000025416.66632.84 (2004).
17. Evans, M. J., Clemens, G., Casey, C. & Baker, M. J. Developing a mobile app for remote access to and data analysis of spectra. *Vib. Spectrosc.* **72**, 37–43, doi: 10.1016/j.vibspec.2014.02.008 (2014).
18. Hughes, C. & Baker, M. J. Can mid-infrared biomedical spectroscopy of cells, fluids and tissue aid improvements in cancer survival? A patient paradigm. *Analyst*, doi: 10.1039/c5an01858g (2015).

## Acknowledgements

The authors would like to thank the EPSRC funded Clinical Infrared and Raman Spectroscopy (CLIRSPEC) Network (EP/L012952/1) for funding and the support of Brain Tumour North West, The Sydney Driscoll Neuroscience Foundation and the Rosemere Cancer Foundation.

## Author Contributions

All authors reviewed the manuscript. C.H. wrote the main manuscript text, supporting information and prepared figures. G.C. contributed to the text, performed sample preparation and acquired data. B.B. contributed to the text, served as technical advisor and acquired data. T.D., K.A., M.J. and A.B. served as clinical advisors. M.W. and E.F. developed the instrumentation for discrete frequency sparse data collection. M.B. and J.R. served as technical advisors. M.J.B. contributed to the text; served as a technical advisor and project manager.

## Additional Information

**Supplementary information** accompanies this paper at <http://www.nature.com/srep>

**Competing financial interests:** These authors declare no competing financial interests.

**How to cite this article:** Hughes, C. *et al.* Introducing Discrete Frequency Infrared Technology for High-Throughput Biofluid Screening. *Sci. Rep.* **6**, 20173; doi: 10.1038/srep20173 (2016).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>