Journal of Nuclear Medicine, published on January 26, 2017 as doi:10.2967/jnumed.116.182311

Title: Evaluation of precision in optoacoustic tomography for preclinical imaging in living subjects.

Author Names and Affiliations:

James Joseph, Michal R Tomaszewski, Isabel Quiros-Gonzalez, Judith Weber, Joanna Brunker and Sarah E Bohndiek*

Department of Physics and Cancer Research UK Cambridge Institute, University of Cambridge, U.K.

Corresponding Author Information*: *Address*: Department of Physics, Cavendish Laboratory, JJ Thomson Avenue, Cambridge, CB3 0HE, U.K. and Cancer Research UK Cambridge Institute, Li Ka Shing Centre, Robinson Way, Cambridge, CB2 0RE, U.K.. Phone: +44 1223 337267; Fax: +44 1223 337000. seb53@cam.ac.uk.

First Author Information: Fellow in training. *Address/phone/fax* as above. jj412@cam.ac.uk

Word Count for Revised Manuscript: 6,380

Financial support for the work: This work was funded by: the EPSRC-CRUK Cancer Imaging Centre in Cambridge and Manchester (C197/A16465); CRUK (C14303/A17197, C47594/A16267); EU-FP7-agreement FP7-PEOPLE-2013-CIG-630729; and the University of Cambridge EPSRC Impact Acceleration Account.

Disclosure: J.J. and M.T. received conference travel funding from iThera Medical.

Running Title: Precision of optoacoustic tomography.

ABSTRACT

Optoacoustic Tomography (OT) is now widely used in preclinical imaging, however, precision (repeatability and reproducibility) of OT has yet to be determined.

Methods: We used a commercial small animal OT system. Measurements in stable phantoms were used to independently assess the impact of system variables on precision (using coefficient of variation, COV), including acquisition wavelength, rotational position, frame averaging. Variables due to animal handling and physiology, such as anatomical placement and anesthesia conditions were then assessed in healthy nude mice using the left kidney and spleen as reference organs. Temporal variation was assessed by repeated measurements over hours and days both in phantoms and *in vivo*. Sensitivity to small molecule dyes was determined in phantoms and *in vivo*; precision was assessed *in vivo* using IRDye800CW.

Results: OT COV in a stable phantom was less than 2% across all wavelengths over 30 days. The factors with greatest impact on the signal repeatability in phantoms were rotational position and user experience, both of which still resulted in a COV of less than 4%. Anatomical region of interest size showed the highest variation at 12% and 18% COV in the kidney and spleen respectively, however, functional SO_2 measurements based on a standard operating procedure showed exceptional reproducibility of <4% COV. COV for repeated injections of IRDye800CW was 6.6%. Sources of variability for *in vivo* data included respiration rate, user experience and animal placement.

Conclusions: Data acquired with our small animal OT system was highly repeatable and reproducible across subjects and over time. Therefore, longitudinal OT studies may be performed with high confidence when our standard operating procedure is followed.

Keywords: Optoacoustic Imaging; Phantoms; *In vivo* imaging; Repeatability; Reproducibility.

INTRODUCTION

Optoacoustic Tomography (OT) is a fast developing imaging modality, combining the high resolution and penetration depth of ultrasound detection with the high contrast available from optical absorption in tissue (*1*). Endogenous optical absorption provides structural and functional imaging information (*2,3*), while the introduction of exogenous contrast agents *in vivo* enables molecular imaging (*4,5*) for example, by targeting specific cell receptors (*6*) or enzymatic processes (*7*). Given that OT is finding widespread application in preclinical imaging (*8*) and has potential for rapid clinical translation (*1*), technical validation assessing both accuracy and precision (*9,10*) is now vital.

To date, technical validation of OT has focused on accuracy, using light fluence models to enable quantitative determination of absorption coefficients (*11*). There is a paucity of data on precision, which can considered as measurement repeatability, with the same subject or user, and reproducibility, with independent subjects or users. Precision is often characterized by the coefficient of variation (COV), which has been assessed to a limited extent for OT in phantoms (*12*), but not *in vivo*, where animal handling and physiology significantly impact results (*8*). High intra-subject COV limits the applicability of a modality for longitudinal imaging studies, whereas high inter-subject COV would modify study design (e.g. requiring that each animal serve as their own control). Here, we developed and applied a framework for evaluating precision of OT systems in phantoms and living subjects. Our results show that precision of the OT system tested compares favorably with existing preclinical imaging modalities and hence could be reliably applied in biomedical research.

MATERIALS AND METHODS

Optoacoustic tomography system

A number of commercial small animal optoacoustic systems are now available (*13*). In this study, we used a MultiSpectral Optoacoustic Tomography (MSOT) system (inVision256-TF; iThera Medical) described in detail elsewhere (*14,15*). Briefly, a tunable (660–1300nm) optical parametric oscillator pumped by an Nd:YAG laser provides 9ns excitation pulses at 10Hz repetition rate. Ten arms of a fiber bundle illuminate a ring of ~8mm width at the sample, with fluence below the maximum permissible exposure according to ANSI Z136.1 over the wavelength range investigated (Supplemental Fig. 1). Laser energy is recorded for each pulse and the acoustic signal is divided by this value prior to saving. The sample is mounted in a motorized holder for linear translation in the z-direction over a range of <150mm (Fig. 1A). Coupling of the sample to the transducers is achieved using a water bath, filled with degassed, deionized water. For ultrasound detection, 256 toroidally focused ultrasound transducers specified at 5MHz center frequency, 60% bandwidth, are organized in a concave array with 270 degree angular coverage and a radius of curvature of 4 cm.

Framework for evaluating precision of preclinical optoacoustic tomography

For preclinical imaging in small animals, evaluation of precision should be made separately across anatomical (e.g. tumor volume), functional (e.g. oxygenation) and molecular (e.g. tracer uptake) parameters. Firstly, to account for system variation, the precision of the imaging data in phantoms should be assessed across all acquisition variables (e.g. wavelengths, rotational positions, averaging, time). Secondly, the precision across living subjects should be assessed while following a standard operating procedure (SOP). Measurements should be taken by different users to evaluate the

impact of user experience and animal handling. Finally, for contrast agent studies, sensitivity to the contrast agent of interest should be determined in addition to precision, to ensure accurate decisions on the dose of agent to be used *in vivo*. Details of standard operating procedures for all experiments presented are given in the Supplemental Note.

Precision evaluation in phantoms and in vivo. A commercial stable polyurethane phantom (fabricated by Computerized Imaging Reference Systems Inc., supplied by iThera Medical) with a diameter of 2cm and sound speed of 1430ms^{-1} was employed for phantom studies of repeatability. The acoustic attenuation of the phantom is given as 0.5 dB/cm/MHz. Using a commercial stable phantom avoids drift in measurements due to matrix degradation and enables users of other systems to replicate our study. The phantom bulk is composed of a purely scattering matrix, within which are contained absorbing calibration targets. The phantom was clamped into the supplied rigid phantom holder, which ensures repeatable positioning, and placed into the imaging chamber of the MSOT system (Fig. 1B). Following equilibration of phantom temperature with the 36°C water bath for 10min, imaging was performed over 5 scan positions covering the calibration targets in 1mm steps. Images were acquired at 660nm and from 700–1100nm in 50nm steps. The repeatability in longitudinal studies was measured over the course of 6 hours, 1 day, and 1 month. Data is shown at 700nm in all cases, with other wavelengths detailed in Supplemental Table 1.

We used the same phantom to investigate the effects of sample rotation and frame averaging. Sample rotation was assessed due to the incomplete acoustic sampling of the system geometry. We acquired data as above with the phantom rotated by 0, 90, 180 and 270 degrees with respect to its long axis between measurements. To test the impact of frame averaging, we recorded up to 100 frames and averaged either the reconstructed image frames (sequential) or the raw acoustic signals (continuous),

available as options within the acquisition software. Given the use of a rigid holder for phantom imaging, the effect of different operators on phantom data acquisition was not explored.

Procedures in small animals were carried out under the authority of project and personal licenses issued by the Home Office, UK and were approved by local Animal Welfare and Ethical Review Bodies. Healthy BALB/c nude mice (Charles River) aged between 8 and 12 weeks were housed in techniplast green line individually ventilated cages with APB6 bedding on a 12 hours on/off light/dark cycle (7am to 7pm) with 5R58 diet.

Mice were anesthetized using inhaled isoflurane mixed with 100% oxygen and placed supine into the animal holder (Fig. 1B). A flexible polyethylene membrane surrounded the mouse forming a watertight seal, with anesthesia supplied via a gas inlet. A thin layer of clear ultrasound gel (Aquasonics, Parker) applied with a spatula was used to couple the skin of the mouse to the membrane. Mouse preparation took ~15min. Ultrasound gel was centrifuged to remove air bubbles and warmed prior to application. Following equilibration of mouse temperature with the 36°C water bath for 12min (previously determined by rectal probe), imaging was performed in a single slice centered on the kidneys and spleen. Images were acquired between 690nm and 880nm in 10nm steps, with 10 averages (continuous averaging).

The influence of inhaled isoflurane concentration (1.2–3%) on respiration rate (40–140 bpm) and the resulting SO_2^{MSOT} measurements was first determined in a subset of animals (n=3), after which respiration was maintained at 60-70bpm with (1.75±0.25)% isoflurane for the remainder of the studies. For full COV calculations, we then examined: continuous imaging over 90min; acquisition of 6 image frames leaving the mouse in the

same position (without replacement, n=7 mice) or removing and immediately replacing the mouse before each acquisition (with replacement, subset of n=4 mice); and repeated imaging over 3 days (with replacement). Given that variability is expected in the handling and placement of the mice, data acquired by two independent users was also analyzed.

Sensitivity and precision for detecting small molecule dyes in phantoms and in vivo. To assess the detection sensitivity for small molecule dyes, phantoms with defined optical properties that closely mimic biological tissue were fabricated. Phantom absorption and reduced scattering coefficients followed the 'generic tissue' definition given by Jacques (*16*) and are shown in Supplemental Figure 2. All chemicals were purchased from Sigma Aldrich unless otherwise stated. Pre-warmed intralipid was added to liquid agar to provide scattering and nigrosin dye to provide absorption. The solution was poured into a 20mL syringe (2cm diameter) containing a 3D printed plastic mold to create a cylindrical hole of 3mm diameter at the phantom center, into which a sealed thin walled plastic straw containing a solution of dye was inserted. Commercially available small molecule dyes (relevant to clinical use) were used: IRDye 800CW (IR800, Licor), indocyanine green (ICG) and methylene blue (MB). IR800 and MB were dissolved in phosphate buffered saline while ICG was dissolved in water due to the poor solubility in phosphate buffered saline. The tissue mimicking phantom was placed into the MSOT system using the procedure described for small animal imaging. The water bath temperature was set to 34°C to prevent loss of structural integrity of the agar. Images were acquired at multiple wavelengths at 5 scan positions with a 1mm step size.

To assess *in vivo* detection sensitivity and precision, we prepared mice (n=3) for imaging (as above) but also inserted a tail vein catheter with a dead volume of 35μL. Under terminal anesthesia (n=2), we performed a dose escalation study, injecting 135μL of 2, 4, 8, 20 and 50μM IR800 dissolved in phosphate buffered saline at 11 min intervals.

In a separate animal, we then repeatedly injected 135µL of 8µM IR800 at 16 min intervals. The narrow time intervals between injections mean that the dye will not have fully cleared from the bloodstream; we therefore calculated the intensity difference before and after injection, rather than comparing absolute intensities. Images were acquired continuously (at the dye specific excitation wavelengths) in order to observe dynamics *in vivo*.

Image and statistical analysis

Images were reconstructed offline and analyzed using the ViewMSOT software package (v3.6; iThera Medical). The model-linear based reconstruction and linear regression multispectral processing tools were used (*17,18*). The reconstruction algorithm includes by default an electrical impulse response correction, which accounts for the Gaussian response of the ultrasound transducers by deconvolving a reference impulse response from the measured acoustic signals. The reference impulse response is based on the vendor specifications for transducer performance, the key parameters of which (5MHz center frequency, 60% bandwidth) were verified by the manufacturer of the supplied transducer array using pulse-echo ultrasound. For the phantom rotation studies, images were reconstructed with and without the impulse response. The same speed of sound (adjusted manually) was used for all image reconstructions within a given data set.

Mean pixel intensity and standard deviation values were extracted from regions of interest (ROIs). For stable phantom studies, the ROI size (12.3mm^2) and position were identical across all data sets (Fig. 1C) and single wavelength images were used for analysis. For tissue mimicking phantoms, the ROI size (6.15mm^2) was identical across all data sets but the position was optimized between images; both single wavelength and

multispectral images were used. For small animal studies, ROIs were drawn around the left kidney and the spleen (Fig. 1C) in each mouse. When IR800 was injected, ROIs were drawn around the right kidney, to avoid any confounding influence of light attenuation by the spleen.

Linear regression with published spectra for oxy- (HbO_2) and deoxy-hemoglobin (Hb), as well as IR800 (*19,20*), was used to produce images in arbitrary units that indicate the relative weight that each spectrum contributes to a given image pixel. Oxygen saturation was extracted as the ratio of ROI data from the HbO_2- and Hb-weighted images and calculated as $HbO_2/(Hb+HbO_2)$. OT is only able to accurately resolve absolute SO_2 if the recorded signal can be related to the absorbed optical energy distribution, which requires knowledge of the light fluence distribution, system response and Grueneisen parameter (*11*); as we do not have this information, our image intensities are given in arbitrary units and we denote the oxygenation metric derived in this study as SO_2^{MSOT}. Total hemoglobin (THb) was extracted as $(Hb+HbO_2)$, hence is given in arbitrary units.

COVs were calculated from raw data extracted from ROIs as the ratio of the standard deviation to the mean, expressed as a percentage. Signal-to-noise ratio (SNR) was calculated as the ratio of the signal ROI mean to the background ROI standard deviation. Signal-to-background ratio was calculated as the ratio of the signal ROI mean to the background ROI mean. Uncertainty on mean values is represented by the standard error unless otherwise stated. Statistical analysis was performed in Origin (OriginLab Corp.).

RESULTS

Precision evaluation in phantoms

The temporal repeatability of the optoacoustic tomography system across all wavelengths assessed with a commercial stable polyurethane phantom was excellent (Supplemental Table 1). No significant drift was seen in the mean pixel intensity during continued acquisition over 160min (Fig. 2A), nor over the course of 6h (Fig. 2B). The COV at 700nm over 6h was 0.5% without replacement (phantom remaining in the system for the full 6h), rising to 1.2% with replacement (phantom removed between data acquisitions). Over 30 days (with replacement, Fig. 2C) we found a COV of just 1.9% with a minor system drift of <0.2%. The laser energy for all short term repeatability studies was stable (Supplemental Fig. 3A-C). A decrease in energy was observed over 30 days (Supplemental Fig. 3D), as would be expected from gradual contamination of the optics, in particular the optical parametric oscillator. When a user with no ROI drawing experience was asked to conduct the task (compared to 2 years experience for the main user), COV for 6h rose to 1.3% and 3.8% without and with replacement respectively and COV over 30 days rose to 3.1%.

We also established the influence of sample rotational position and image averaging. The 0° and 180° rotations are located within similar view fields of the transducer array, whereas the 90° rotation is likely to be less well captured than the 270° rotation due to the limited bandwidth and field of view of the ultrasound transducers, as confirmed by images reconstructed without the correction (Fig. 3A). Electrical impulse response correction of the raw data reduced the COV due to rotation from 13.4% (Fig. 3B) to 3.6% (Fig. 3C) and almost doubled the image SNR (from 46.1 to 91.5). Although our transducer array meets design specifications within tolerances, the transducers that

capture the left hand side of the image have a tendency towards higher center frequency and bandwidth (manufacturer certificate of conformance). This results in an overamplification of the 0° rotation compared to 180° after correction (Fig. 3C). The improved representation of low frequencies (where the field of view of the individual transducers is much wider) results in an overall increase in the mean pixel intensity after correction and more comparable data at the 90° and 270° rotations.

For image averaging, the sequential method (averaging of multiple reconstructed image frames, Supplemental Fig. 4A) increased SNR but continuous (direct averaging of the acoustic signals, Supplemental Fig. 4B) did not. This is likely due to the fact that the acoustic signal averaging smooths the raw pressure signals so does not account for all factors that influence the noise in the final reconstructed image. These data illustrate the trade-off between SNR, which was optimal with >25 averages (Supplemental Fig. 4A), and acquisition time, which was >25s per data set with this number of averages.

As a preliminary assessment of the generalizability of these precision measurements, we also acquired images of the commercial phantom using a non-tomographic optoacoustic system, which uses a light emitting diode array for signal excitation and a linear ultrasound transducer array for detection (Supplemental Fig. 5A). The acquired data indicated that the same commercial phantom could be successfully used in a different geometry. Measured COVs were 13.9% and 6.1% over 160min and 6h respectively (Supplemental Fig. 5B,C); these higher values relative to the OT system could be due to heating of the excitation light source and water coupling medium over time.

Precision evaluation in vivo

Given the impact of animal handling and physiology on *in vivo* imaging data, we established the precision of our OT system in living subjects using the left kidney and spleen as reference organs. All ROIs were drawn by an experienced (>1 year) user unless otherwise stated. Considering first anatomical size data, the left kidney and spleen had average ROI areas of (21.0±0.7)mm^2 (COV 12%) and (19.2±1.0)mm^2 (COV 18%) respectively over all mice (n=13). When ROIs were independently drawn in repeated frames of individual mice, the average COVs were 4.4±1.2% and 7.3±0.7% for the kidney and spleen (n=8), increasing to 6.1±1.1% and 10.4±1.5% respectively with replacement between acquisitions (n=7). For the following studies we used a fixed ROI size and shape for each mouse, copying into repeated image frames and allowing adjustments in orientation to position the ROI on the organ.

Having established the precision of anatomical size data, we then moved on to functional data. Example images of Hb and HbO_2 are shown in Supplemental Fig. 6. We noted that ROI data extracted from the Hb and HbO_2 images directly, as well as their sum (THb) showed a high COV <18.5% between mice (reproducibility); the ratiometric readout SO_2^{MSOT} showed a much lower COV with <4% variation (Table 1). We attributed the high variation in Hb, HbO_2 and THb metrics to inherent biological variation between mice, which appears to be minimized by taking the ratio, hence we used SO_2^{MSOT} as our OT biomarker for temporal COV calculations. We also noted that changing the isoflurane anesthetic concentration in the range 1.2–3% over 10mins (associated respiration rate range 40–140 bpm) resulted in >30% decrease in spleen SO_2^{MSOT} and >9% decrease in the kidney (p=0.047, 0.043 respectively by paired two-tailed t-test; n=3). For the temporal studies, we therefore maintained all mice using isoflurane anesthetic concentration (1.75±0.25)% (respiration rate range 60-70bpm) during imaging to minimize variation.

Temporal repeatability *in vivo* was assessed over 90mins (one continuous acquisition) then during 6 consecutive acquisitions on one day (with and without replacement) and over 3 days. COVs are calculated for the ratiometric SO_2^{MSOT} measurement; THb is shown for comparison. We observed a drift in SO_2^{MSOT} and THb over 90mins (Fig. 4A), but for repeated measurements, we found excellent longitudinal repeatability over 6 acquisitions within the same day (Fig. 4B). While this remained true for SO_2^{MSOT} over 3 days, it was not the case for THb (Fig. 4C). Since the animal holder is not rigid, we examined the impact of replacement and also of different users. Replacement of the mouse between sequential acquisitions led to a COV of <1.4% compared to <0.6% without replacement (Table 2). A less experienced user (acquisition and ROI drawing performed in fewer than 10 mice) resulted in only a marginally higher COV (<0.8%). One-way repeated measures ANOVA performed for the left kidney and spleen, SO_2^{MSOT} and THb, indicated that replacements had no significant influence on the variation ($p>0.27$) but individuals were significantly different ($p<0.0001$). Over 3 days, neither replacements ($p>0.09$) nor individuals ($p>0.47$) were significant. Overall COV in SO_2^{MSOT} was <1% for repeatability and <4% for reproducibility.

Sensitivity for small molecule dyes in phantoms and in vivo

Detection limits were determined to be 100nM for IR800 (Supplemental Fig. 7A) and ICG (Supplemental Fig. 7B) and 5μM for MB (Supplemental Fig. 7C); equivalent optical density values were 0.10 (IR800, 777nm), 0.084 (ICG, 778 nm) and 0.17 (MB, 664 nm) respectively. Considering the volume of dye present within an image slice (determined by the transducer focus as approximately 6.16μL), this is equivalent to ~0.62pmol for IR800 and ICG, and ~30.8pmol for MB. The higher threshold for MB may be explained by the smaller extinction coefficient higher background absorption at its peak wavelength. The absorption spectra (Supplemental Fig. 7) recovered from the

multiwavelength optoacoustic data generally follow those acquired independently on a UV-Vis plate reader, although some discrepancies are observed for IR800 and ICG at shorter wavelengths.

We tested *in vivo* sensitivity and precision for IR800 (Fig. 5). We first performed dose escalation studies (n=2). We started injections from 2μM based on the detection limits derived above and assuming ~1:25 dilution of the injected contrast agent in the blood stream of the mouse (~100μL injection in ~2.5mL blood pool). As expected, there was no enhancement at 2μM, only at 4μM. A clear kinetic curve was apparent for 8μM (Fig. 5A), which also gave the best reproducibility in this limited study (Fig. 5B). For repeated 8μM injections of in one mouse, a COV of 6.6% was derived for the rise above baseline after each injection.

DISCUSSION

Here, we developed and applied a framework that can be used for evaluating the precision of OT in phantoms and *in vivo* (*21,22*). Data acquired in a stable phantom showed exceptional repeatability of the OT system tested, with COV <2% across all wavelengths over 30 days. Across all parameters tested, the factors that demonstrated greatest impact on the signal repeatability were rotational position and ROI drawing by an inexperienced user, which nonetheless produced a COV of less than 4%.

We explored both repeatability (same subject) and reproducibility (between subjects) *in vivo*. For anatomical data, computed tomography and magnetic resonance imaging typically show <5% COV for measurements of ROI size both within and between animals (*23–25*). We found relatively high COVs of 12% and 18% in the kidney and spleen ROI sizes respectively, when compared across all subjects. However, when data were extracted from repeated ROI drawing within the same mouse (with replacement), the values fell to 6% and 10% respectively. Our findings indicate that the primary source of variation in size is anatomical positioning within the non-rigid animal holder. In addition, as contrast is defined by the concentration of blood vasculature rather than a clear organ boundary, ROI drawing for OT is likely to be more subjective than for a modality like CT.

We observed a significant drift in functional parameters SO_2^{MSOT} and THb over 90mins. It has been previously documented that anesthesia induces up to 20% change in blood mean hemoglobin concentration (*26*) and that isoflurane can also suppress the erythropoietin hypoxia response and cause vasodilation (*27,28*), which may in part account for the observed drift. The larger effect observed in the spleen may be due to the increased hemoglobin concentration of splenic blood, which would accentuate these

effects (*29*). The use of water submersion for acoustic coupling may also influence our results, although to the best of our knowledge, the influence of external hydrodynamic pressure in mouse hemodynamics has not been quantified. These factors may also play in to the longer term drift observed in THb over 3 days; further work is necessary to understand the source of this variation. Nonetheless, in our SOP we controlled as far as possible for short-term temporal drift by imaging at a fixed time point, and for the impact of isoflurane concentration (and associated respiration rate) by maintaining consistent values across imaging sessions. Using OT to derive SO_2^{MSOT} values with our SOP, we found that the COV across mice was always <4% and within the same mouse was <2%. For functional oxygenation data extracted from blood oxygen level dependent magnetic resonance imaging or arterial spin labeling measurements, COVs in the range of 10-15% are often reported (*30,31*). The relatively low COV values measured in our OT system indicate that it could be a competitive approach for preclinical functional imaging.

Finally we examined dye sensitivity and precision in tissue mimicking phantoms and *in vivo*. For IR800, the detection sensitivity limit in phantoms of 100nM translated directly into limiting *in vivo* dose of 4μM based on the expected dilution of the injected volume in the blood stream. Performing dye sensitivity studies in tissue mimicking phantoms can therefore assist selection of an appropriate dose for *in vivo* studies for untargeted contrast agents. The COV of the signal rise within the kidney over repeated injections was 6.6%. For dynamic contrast enhanced magnetic resonance imaging, which uses injected gadolinium to assess perfusion and permeability, COV in the derived metric K_{trans} is typically greater than 10% (*24,32*). Again, our OT system therefore appears competitive for preclinical functional imaging.

There remain some limitations to this study, which was conducted in a single center using equipment from a single vendor. In future, multi-center studies across

multiple vendors would be needed to assess overall precision of OT. It would also be prudent to perform tests over a range of common mouse strains with independent physiological measurement, since hemodynamics and skin pigmentation vary significantly between strains. For the dye studies, we used a limited number of mice and an untargeted contrast agent, which cannot provide insight into the dose required for molecular imaging studies where circulatory and cellular barriers must be overcome. Such studies should be expanded in future to test targeted contrast agents, assessing molecular imaging precision compared to other modalities, for example, positron emission tomography (*33*) and hyperpolarized ^{13}C magnetic resonance imaging (*34*). Future comparison of anatomical, functional and molecular imaging measurements made with OT to established preclinical imaging approaches will help to establish acceptable performance criteria for OT metrics.

Nevertheless, we are able to make a number of recommendations based on our findings for future studies *in vivo* using the OT system tested. Firstly, the influence of anatomical variation, including biological differences and physical (rotational/organ) position on functional metrics can be minimized by extracting the ratiometric parameter SO_2^{MSOT}. Secondly, both data acquisition and ROI analysis for a single study should be performed by users with the same level of experience. Finally, an SOP with consistent animal handling and physiology, including a fixed duration and concentration range of anesthesia (resulting in stable respiration rate), minimizes variation between mice. If these recommendations are followed, the OT system tested provides a high precision readout for preclinical imaging.

CONCLUSION

We investigated OT precision across key variables including time and user experience in phantoms and *in vivo*. We find that the optoacoustic tomography system tested shows performance consistent with or exceeding other reported values for preclinical imaging modalities, highlighting the potential for widespread acceptance as a routine tool in preclinical biomedical research in the near future.

ACKNOWLEDGMENTS

We would like to thank all at iThera Medical (especially Stefan Morscher) for helpful discussions and provision of ultrasound transducer data sheets. We are especially grateful to Dr Naoto Sato and Dr Toshitaka Agano with all at PreXion for the loan of their light emitting diode based optoacoustic imaging system. We also thank the CRUK CI BRU and Dr James Mason of the MRC BSU for their technical assistance in the completion of this study.

REFERENCES

1. Taruttis A, Van Dam GM, Ntziachristos V. Mesoscopic and macroscopic optoacoustic imaging of cancer. *Cancer Res*. 2015;75:1548-1559.

2. Guggenheim JA, Allen TJ, Plumb A, et al. Photoacoustic imaging of human lymph nodes with endogenous lipid and hemoglobin contrast. *J Biomed Opt*. 2015;20:50504.

3. Bohndiek SE, Sasportas LS, Machtaler S, Jokerst J V, Hori S, Gambhir SS. Photoacoustic tomography detects early vessel regression and normalization during ovarian tumor response to the antiangiogenic therapy trebananib. *J Nucl Med*. 2015;56:1942-1947.

4. James ML, Gambhir SS. A molecular imaging primer: modalities, imaging agents, and applications. *Physiol Rev*. 2012;92:897-965.

5. Weber J, Beard PC, Bohndiek SE. Contrast agents for photoacoustic imaging. *Nat Methods*. 2016;13:639-650.

6. Hudson S V., Huang JS, Yin W, et al. Targeted noninvasive imaging of EGFR-expressing orthotopic pancreatic cancer using multispectral optoacoustic tomography. *Cancer Res*. 2014;74:6271-6279.

7. Levi J, Kothapalli S-R, Bohndiek S, et al. Molecular photoacoustic imaging of follicular thyroid carcinoma. *Clin Cancer Res*. 2013;19:1494-1502.

8. Stout D, Berr SS, LeBlanc A, et al. Guidance for methods descriptions used in preclinical imaging papers. *Mol Imaging*. 2013;12:1-15.

9. European Society of Radiology (ESR). *ESR Statement on the Stepwise Development of Imaging Biomarkers*. Vol 4.; 2013.

10. O'Connor JPB, Aboagye EO, Adams JE et al. Imaging biomarker roadmap for cancer studies. *Nat Rev Clin Onc.* 2016; Advance Online Publication.

11. Cox B, Laufer JG, Arridge SR, Beard PC. Quantitative spectroscopic photoacoustic imaging: a review. *J Biomed Opt*. 2012;17:61202.

12. Bohndiek SE, Bodapati S, Van De Sompel D, Kothapalli SR, Gambhir SS. Development and application of stable phantoms for the evaluation of photoacoustic imaging instruments. *PLoS One*. 2013;8:1-14.

13. Optoacoustic Tomography Vendors at 4/2016. www.visualsonics.com/products/vevo-lazr-photoacoustic-imaging; www.tomowave.com; www.endrainc.com; www.pst-inc.com; www.ithera-medical.com.

14. Morscher S, Driessen WHP, Claussen J, Burton NC. Semi-quantitative multispectral optoacoustic tomography (MSOT) for volumetric PK imaging of gastric emptying. *Photoacoustics*. 2014;2:103-110.

15. Dima A, Burton NC, Ntziachristos V. Multispectral optoacoustic tomography at 64,

128, and 256 channels. *J Biomed Opt*. 2014;19:36021.

16. Jacques SL. Optical properties of biological tissues: a review. *Phys Med Biol*. 2013;58:R37-61.

17. Dean-Ben XL, Buehler A, Ntziachristos V, Razansky D. Accurate model-based reconstruction algorithm for three-dimensional optoacoustic tomography. *IEEE Trans Med Imaging*. 2012;31:1922-1928.

18. Tzoumas S, Deliolanis N, Morscher S, Ntziachristos V. Unmixing molecular agents from absorbing tissue in multispectral optoacoustic tomography. *IEEE Trans Med Imaging*. 2014;33:48-60.

19. Prahl SA. Optical Properties Spectra. http://omlc.org/spectra/. Accessed 9/2015.

20. Licor. IRDye® Infrared Dyes. https://www.licor.com/bio/products/reagents/irdye/index.html. Accessed 9/2015.

21. Vanhove C, Bankstahl JP, Krämer SD, Visser E, Belcari N, Vandenberghe S. Accurate molecular imaging of small animals taking into account animal models, handling, anaesthesia, quality control and imaging system performance. *EJNMMI Phys*. 2015;2:31.

22. Hildebrandt IJ, Su H, Weber WA. Anesthesia and other considerations for in vivo imaging of small animals. *ILAR J*. 2008;49:17-26.

23. Umoh JU, Sampaio A V, Welch I, Pitelka V. In vivo micro-CT analysis of bone remodeling in a rat calvarial defect model. *Phys Med Biol*. 2009;54:2147-2161.

24. Subashi E, Choudhury KR, Johnson GA. An analysis of the uncertainty and bias in DCE-MRI measurements using the spoiled gradient-recalled echo pulse sequence. *Med Phys*. 2014;41:32301.

25. Montelius M, Ljungberg M, Horn M, Forssell-aronsson E. Tumour size measurement in a mouse model using high resolution MRI. *BMC Med Imaging*. 2012;12:1-7.

26. Gothelf A, Hojman P, Gehl J. Change in hemoglobin levels due to anesthesia in mice: An important confounder in studies on Hematopoietic drugs. *Biol Proced Online*. 2009;11:325-330.

27. Kai S, Tanaka T, Matsuyama T, Suzuki K, Hirota K. The volatile anesthetic isoflurane differentially suppresses the induction of erythropoietin synthesis elicited by acute anemia and systemic hypoxemia in mice in an hypoxia-inducible factor-2-dependent manner. *Eur J Pharmacol*. 2014;732:43-49.

28. Masamoto K, Fukuda M, Vazquez A, Kim SG. Dose-dependent effect of isoflurane on neurovascular coupling in rat cerebral cortex. *Eur J Neurosci*. 2009;30:242-250.

29. MacDonald IC, Schmidt EE, Groom AC. The high splenic hematocrit: A rheological consequence of red cell flow through the reticular meshwork. *Microvasc Res*. 1991;42:60-76.

30. Alonso BDC, Lowe AS, Dear JP, Lee KC, Steven CR. Sensory inputs from whisking movements modify cortical whisker maps visualized with functional magnetic resonance imaging. *Cereb Cortex*. 2008;18:1314-1325.

31. Su Y, Thomas JB, Hauch AC, Hershey T, Ances BM, Arbela AM. Comparison of Regional Cerebral Blood Flow Responses to Hypoglycemia Using Pulsed Arterial Spin Labeling and Positron Emission Tomography. *PLoS One*. 2013;8:e60085.

32. Aryal MP, Nagaraja T., Brown SL, et al. Intra-tumor distribution and test-retest comparisons of physiological parameters quantified by dynamic contrast-enhanced MRI in rat U251 glioma. *NMR Biomed*. 2008;141:520-529.

33. Martic-Kehl M, Ametamey S, Alf M, Schubiger P, Honer M. Impact of inherent variability and experimental parameters on the reliability of small animal PET data. *EJNMMI Res*. 2012;2:26.

34. Serrao EM, Rodrigues TB, Gallagher FA, et al. Effects of fasting on serial measurements of hyperpolarized [1-13C]pyruvate metabolism in tumors. *NMR Biomed*. 2016;29:1048-1055.

Figure Legends

Figure 1. Illustration of system geometry and sample preparation. The sample is placed within the imaging chamber using specialized phantom or mouse holders. (A) Schematic of the sample holder geometry within the ultrasound transducer ring. (B) Photographs of the two holders and their placement in the system. The resulting images (C) of the stable polyurethane phantom (top) and mouse cross section (bottom) are illustrated with the regions of interest (ROIs) used for signal (including kidney and spleen *in vivo*) and background (bkgd) in the analysis outlined. Scale bar =4mm.

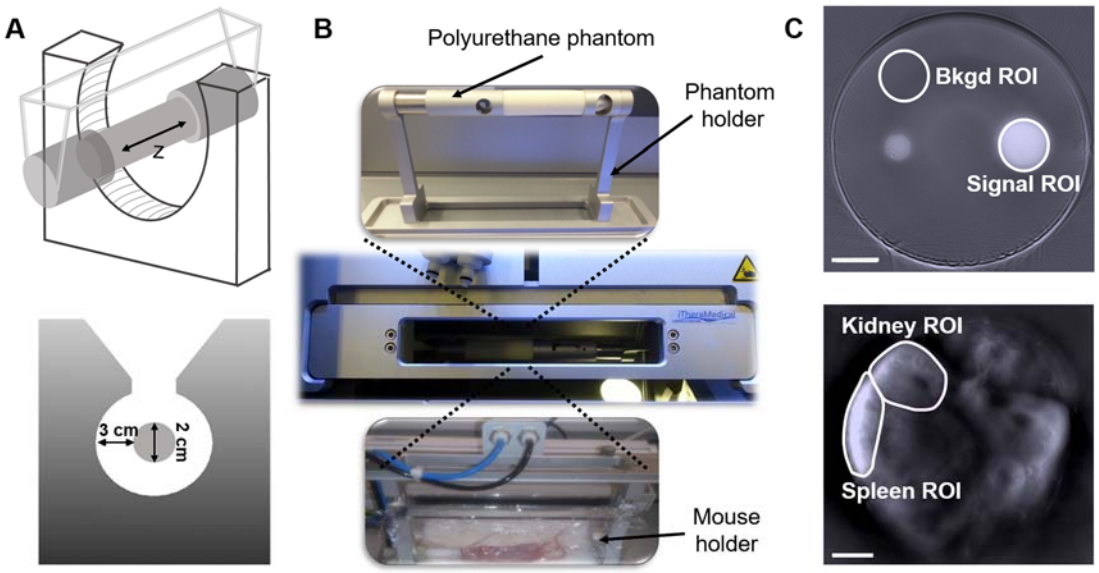


Figure 2. Optoacoustic precision as a function of time in the stable polyurethane phantom at 700nm. (A) Normalized mean pixel intensity over 160mins (NR) with slope of $(-6.34\pm1.62)\times10^{-5}$a.u. (B) Mean pixel intensity (arbitrary units, a.u) over 6h in a single day with replacement (R) and without replacement (NR) of the phantom between data acquisitions; slopes are (-30.7±28.2)a.u. and (-4.9±29.2)a.u. respectively. (C) Mean pixel intensity over 30 days (R) with a slope of (50.4±4.0)a.u., equating to a 0.17% drift. Data were extracted from the signal ROI and represent an average of n=5 scan positions per time point. Replacement indicates removal of the phantom from the imaging system between measurements; average SNR for replacement data was 239±24 and 234±12 for 6h and 30 days respectively. Error bars are within the symbols.

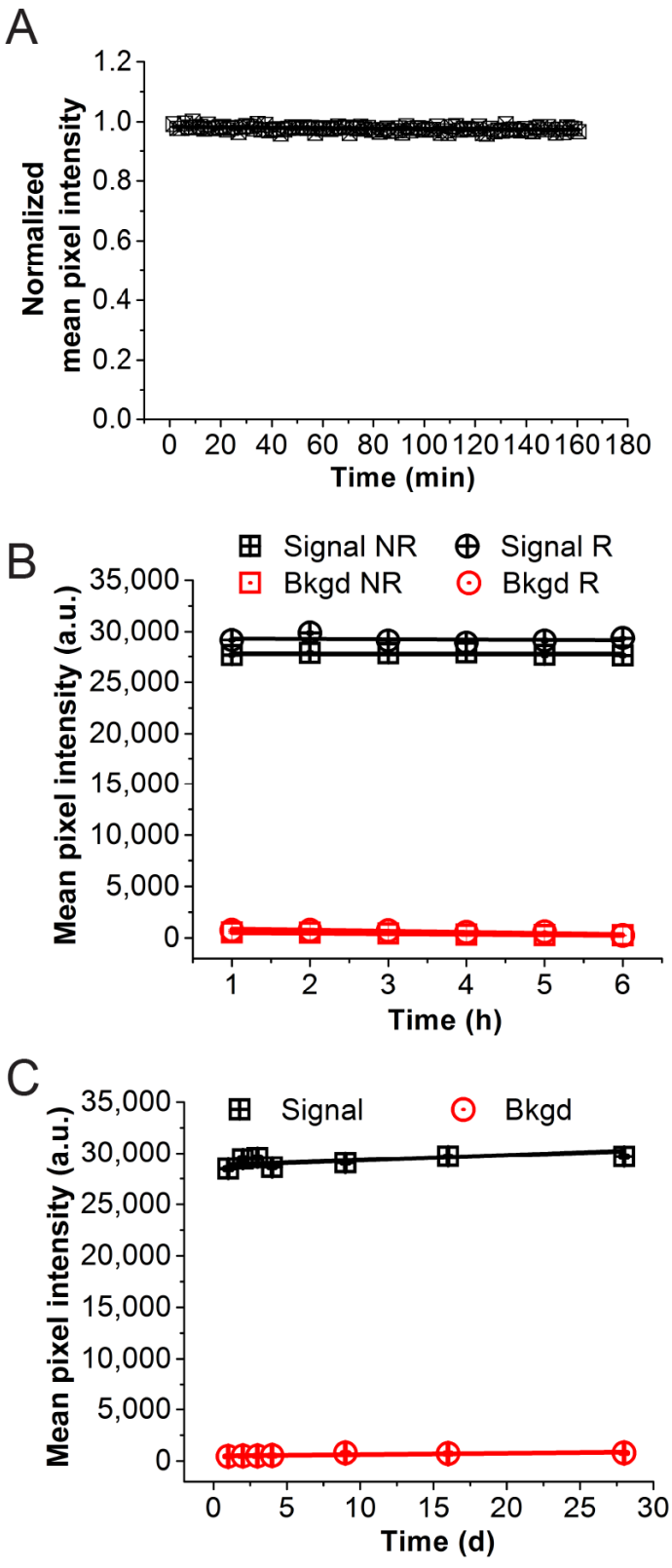

A
Normalized mean pixel intensity
1.2
1.0
0.8
0.6
0.4
0.2
0.0
0 20 40 60 80 100 120 140 160 180
Time (min)
B
Signal NR
Signal R
Bkgd NR
Bkgd R
Mean pixel intensity (a.u.)
35,000
30,000
25,000
20,000
15,000
10,000
5,000
0
1 2 3 4 5 6
Time (h)
C
Signal
Bkgd
Mean pixel intensity (a.u.)
35,000
30,000
25,000
20,000
15,000
10,000
5,000
0
0 5 10 15 20 25 30
Time (d)

Figure 3. Impact of sample orientation within the imaging chamber. (A) Reconstructed images of the stable polyurethane phantom four different rotations (0, 90, 180, 270°) reconstructed without impulse response correction. The white lines indicate the location of the ultrasound transducer array. Mean pixel intensities from an ROI placed over the larger calibration target (signal ROI in Fig. 1C) were extracted without (B) and with (C) impulse response correction. Data in (B,C) represent an average of n=5 scan positions, with error bars showing the standard error on the mean. The SNR (measured across all positions with a background ROI placed at the phantom center) increased after impulse response correction, rising from 46.1 to 91.5. Scale bar =4mm.

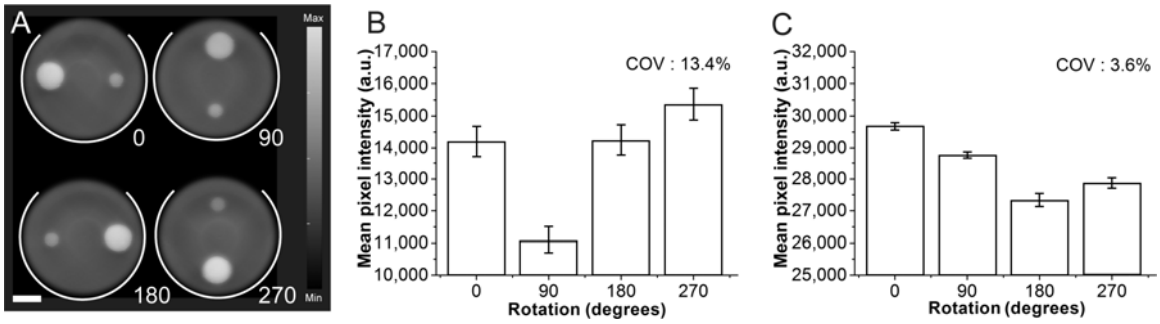


Figure 4. Optoacoustic precision as a function of time *in vivo.* Repeatability for *in vivo* studies was assessed over 90mins (as a representative single scan time), 6 consecutive replacements (to assess influence of positioning) and over 3 days (mimicking a longitudinal study). (A) Oxygen saturation (SO_2^{MSOT}) and total hemoglobin (THb) measured over 90mins [n=3 mice; SO_2^{MSOT} slope=-(9.5±1.9)x10^{-4}%min^{-1} left kidney, -(12±1.4)x10^{-4}%min^{-1} spleen; THb slope=-(30.81±3.05)x10^{-4}a.u.min^{-1} left kidney, -(143.8±11.78)x10^{-3}a.u. min^{-1} spleen]. (B) Over 6 replacements little variation was seen [n=7 mice; SO_2^{MSOT} slope=(30.4±9.95)x10^{-4}%, left kidney, (17.3±6.10)x10^{-4}% spleen; THb slope=(0.02±0.03)a.u. left kidney, (0.13±0.08)a.u. spleen]. (C) SO_2^{MSOT} was also relatively stable over 3 days, but greater variation was seen in THb [n=3 mice; SO_2^{MSOT} slope=(18.9±1.2)x10^{-3}%day^{-1} left kidney, (37.6±27.54)x10^{-3}%day^{-1} spleen; THb slope=(1.2±0.5)a.u.day^{-1} left kidney, (-4.5±0.6) a.u.day^{-1} spleen]. According to regression analysis, linear fits in B (SO_2^{MSOT} and THb) and C (SO_2^{MSOT} only), are not ‘significantly non zero’, indicating that no systematic variation can be observed in these data. Error bars are within the symbols unless visible within the graph.

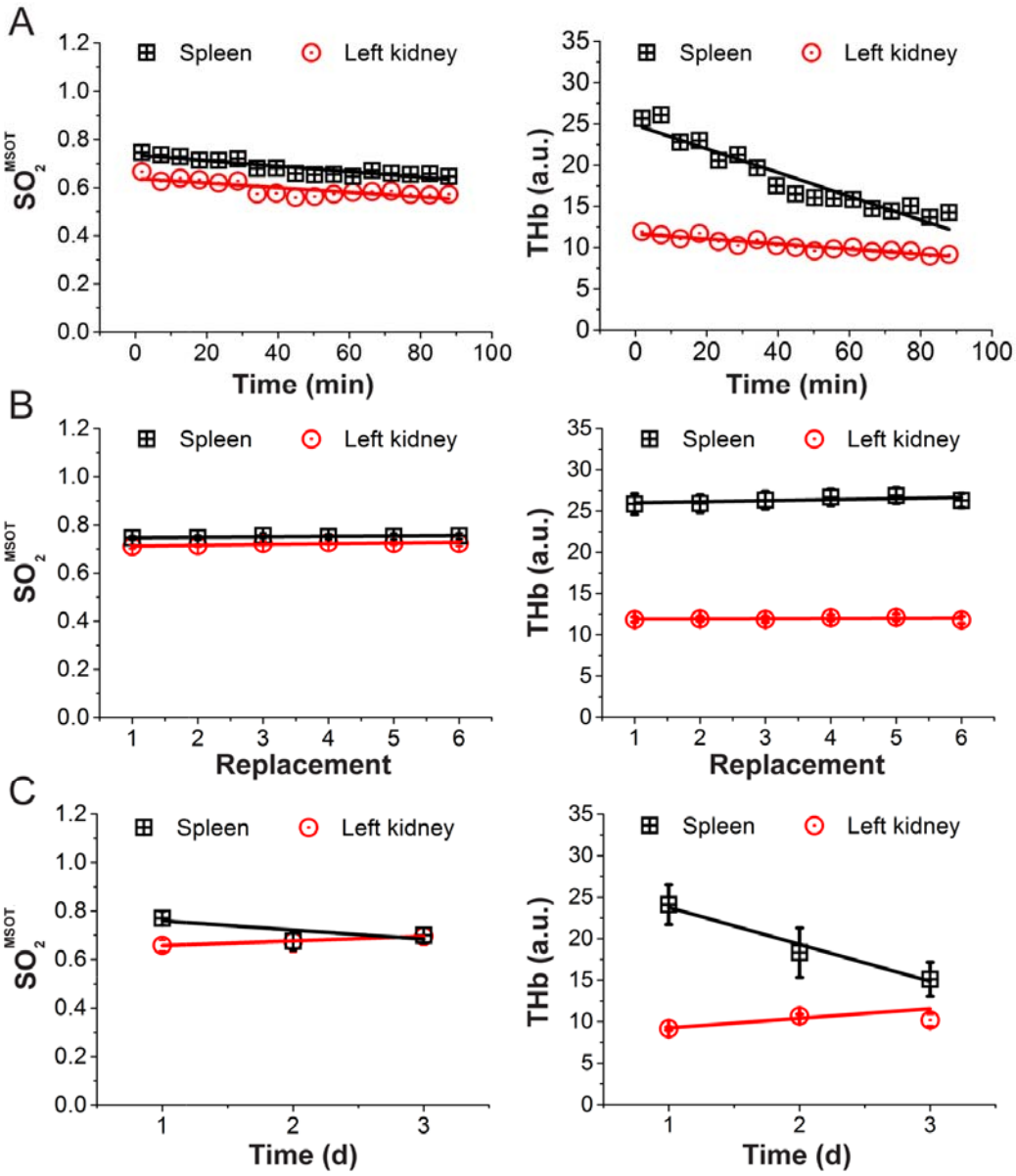
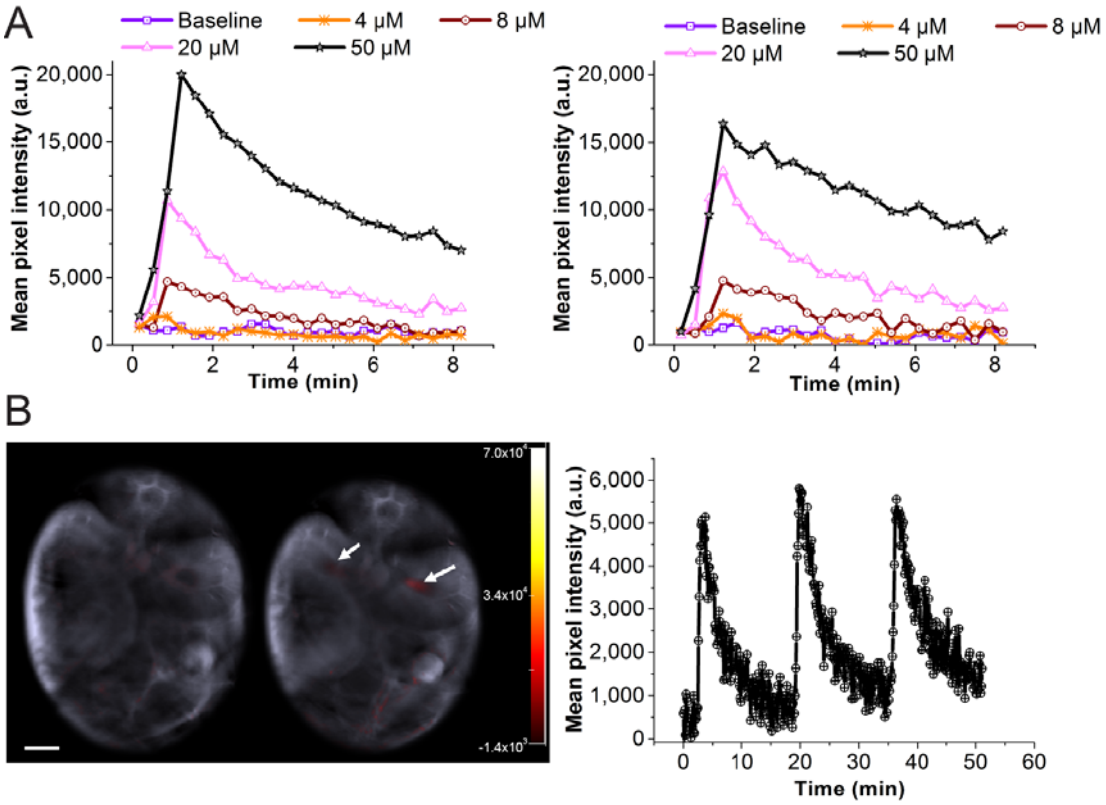

A
Spleen
Left kidney
SO_2^{MSOT}
Time (min)
THb (a.u.)
B
Replacement
C
Time (d)

Figure 5. Sensitivity and precision for *in vivo* imaging of IR800CW. Dose escalation study of injection of 100μL IR800 at concentrations of 4, 8, 20 and 50μM using an ROI drawn on the right kidney of two separate mice (A). Injections were made sequentially at 11min intervals and data are time and intensity shifted to show the point of injection at t=0min. 8μM was identified as ideal for *in vivo* visualization (B), as only a small enhancement was seen at 4μM. Images in (B) are pre-injection (left) and at 1min post-injection (right). White arrows indicate contrast agent inflow within the kidneys; scale bar is 3 mm. Mean pixel intensity for 3 repeated injections at 16min intervals of the same concentration and volume are also shown.



Tables

Table 1: Assessing reproducibility (between mice) using the coefficient of variation (COV) of the oxy- (HbO_2) and deoxyhemoglobin (Hb) signals compared to oxygen saturation (SO_2^{MSOT}**).** COV is of the ROI mean extracted from 7 mice.

	Coefficient of Variation (%)			
Organ ROI	HbO_2	Hb	THb (HbO_2 + Hb)	SO_2^{MSOT} $HbO_2/(HbO_2+Hb)$
Spleen	16.4	13.8	12.9	3.9
Kidney	18.0	18.4	17.2	3.4

Table 2: Assessing the influence of animal preparation and stability in the chamber on the coefficient of variation (COV) of oxygen saturation (SO_2^{MSOT}**).** 'Per mouse, with replacement' is the average COV calculated over 6 image frames acquired per mouse imaged by a single experienced user with replacement i.e. mouse removed from the imaging chamber in between measurements (n=4 refers to the 4 replicates of the experiment). 'Per mouse, no replacement' is the average COV calculated from 6 image frames acquired per mouse without replacement by a single experienced user (n=4) and a second inexperienced user (n=3). The average COV over the total n=7 replicates is also given. Error bars represent the standard error on the mean.

	Coefficient of Variation (%)			
	Per Mouse, With Replacement	Per Mouse No Replacement		
Organ ROI	*(1 experienced user, n=4)*	*(1 experienced user, n=4)*	*(1 inexperienced user, n=3)*	*(2 users, n=7)*
Spleen	1.35±0.34	0.54±0.08	0.79±0.16	0.76±0.29
Kidney	1.02±0.15	0.34±0.05	0.64±0.07	0.47±0.18

Supplementary Note: Standard operating procedures for optoacoustic experiments in phantoms and *in vivo*

Primary sources of variation in optoacoustic image data include hardware instabilities and user operational experience. Hardware instabilities represent the limit of precision of a given optoacoustic system; these are primarily:

1. Light source energy fluctuations, which may arise due to
    a. Lack of temperature stabilization of laser cavity and tuning crystal;
    b. Aging of the pump source;
    c. Dirty or damaged optics;
    d. Occurrence of timing jitters etc.
2. Ambient system noise, which may arise due to
    a. Electronic interference;
    b. Overheating of the control electronics.

User operational experience can impact the precision of a given OT system due to differences in:

1. Coupling efficiency during data acquisition, as the user must ensure that the coupling medium (e.g. water, gel) is free of bubbles and maintained at a similar temperature to the sample.
2. Sample positioning during data acquisition, which can be optimized with the use of motorized translation stages.
3. Region of interest placement during data analysis, which can be minimized by copying regions of interest between data sets.

Intrinsic variations in the sample (e.g. mouse) are frequently the subject of a given study, however, for *in vivo* imaging unwanted variation may also occur as a result of animal handling or physiology, which must be evaluated and accounted for in the given mouse strain, age and gender of interest for an experiment. Data acquisition and

reconstruction parameters are usually treated as variables to be controlled except when they themselves are the subject of the test.

The standard operating procedure presented below has been designed to minimize and quantify the impact of the identified sources of variation.

MATERIALS

Reagents

- Agar (Sigma Aldrich)
- Nigrosin (Sigma Aldrich)
- Intralipid 20% Emulsion (Sigma Aldrich)
- Contrast agents (e.g. Methylene Blue and Cardiogreen, Sigma Aldrich; IR800CW Carboxylate, Licor)
- Gluegun and 3mm polyethylene tubes
- Isoflurane (any vendor)
- Ultrasound gel (Aquasonics, Parker)
- Eye hydration ointment (any vendor)
- Phosphate buffered saline, pH 7.4 (Gibco, Life Technologies)

Equipment

For data acquisition

- Commercial optoacoustic imaging system, including:
    - Tunable pulsed light source;
    - Light delivery optics, which may include fiber bundles, focusing, beam-shaping and beam-sampling optics;
    - Ultrasound transducer array;
    - Motorized translation stages and suitable controllers, if required;

    - Spectrometer and power meter for routine monitoring of (and offline correction for) laser energy output;
    - Data acquisition system with signal conditioning units;
    - Reconstruction and data analysis computer;
    - Instrument control, reconstruction and analysis software.
- Phantom and animal holders to provide repeatable positioning of the sample.

For phantom preparation

- Weighing scales or balance.
- Microwave oven.
- Warm water bath operating at up to 60°C.
- Spectrophotometer with the same wavelength range as the tunable pulsed light source.
- Phantom molds (shape determined by the optoacoustic imaging system geometry).
    - Molds can be easily formed in desired geometry using a 3D printer.
- Polyethylene tubes (with minimum wall thickness and of a suitable diameter for the given optoacoustic imaging system geometry).

For animal experiments

- Inhalation anesthesia delivery system.
- Heating system.
- Venous catheter for intravenous contrast agent administration.
- Polyethylene membrane, to avoid direct contact with animal during imaging procedure.
- Ideally, a physiology monitoring system would be used to assess temperature, respiration and pulse oximetry.

Samples

- Commercial stable polyurethane phantom (fabricated by Computerized Imaging Reference Systems Inc., supplied by iThera Medical)
- Agarose phantoms (see below)
- Healthy BALB/c nude mice (~ 18 g, Charles River). Full details of the housing conditions for the mice are given in the main manuscript.
    - Although immune compromised mice were used for our studies, other strains of mice could be used upon hair removal around the region of interest.
    - Replicates should be performed with the same strain, age and health condition.

PROCEDURES

Tissue mimicking phantom fabrication

Recipe to make 100 mL solution with 0.05 cm^{-1} absorption coefficient and 5 cm^{-1} scattering coefficient

- Weigh 15 mg of nigrosin and add to 30 mL deionised water in a 50 mL falcon tube. Measure the absorption spectrum of the solution using a spectrophotometer and ensure that a 1:2 dilution series in additional deionised water maintains linearity.
- Measure 2.08 mL of intralipid into a 15 mL falcon tube and pre-warm in a water bath at 50 °C for at least 5 minutes.
- Weigh 1.5 g of agar and add to 97.3 mL deionised water in a glass media bottle. With the cap placed loosely on the bottle, heat the solution inside a microwave

oven for around 1 minute (depending on power rating), until it visibly boils and the solution appears clear

    - *Important, do not let the solution overflow the flask.*
- Allow the solution to cool down to less than 60°C and add the pre-warmed intralipid into the agar solution.
- Immediately measure 0.62 mL of the nigrosin solution and mix thoroughly to form a homogenous solution.
- Immediately pour the solution into suitable phantom moulds and leave the solution to set in room temperature or inside a refrigerator
    - *Important, always use the same method for allowing the phantom to set; room temperature or refrigeration may result in different phantom properties.*

Preparing for optoacoustic data acquisition

The optoacoustic imaging system (including light source and tuning units, as well as any coupling water bath) should be switched on sufficiently far in advance of the imaging session to allow adequate warm up time of the given light source, according to the suggestion of the manufacturers. Typical values range from 15-30 minutes. All acquisition hardware and any motorized translation stages should be initialized and checked prior to data acquisition.

Data acquisition variables should be identified for the optoacoustic imaging system under test. In this study, we examined the effects of changing wavelength, rotational sampling positioning and signal averaging. Other variables that could be tested, but were not relevant for our system, include: number of views; ultrasound frequency (if multiple handheld probes are available, for instance); and gain / time gain compensation (particularly in linear transducer arrays). Each variable should be adjusted

in turn and image data recorded from a stable phantom. Once these data have been evaluated, a defined imaging protocol should be set for an entire series of comparable samples (e.g. a longitudinal phantom or in vivo study).

Phantom imaging

Phantoms may be composed of a stable material (such as the commercial polyurethane phantom used in this study, or a photostable material such as a carbon fibre) or a tissue mimicking material (as found in the above recipe). Images should be periodically acquired from a stable phantom to verify that the OT system is meeting performance expectations. Tissue mimicking phantoms are used to evaluate the system sensitivity and precision for detection of contrast agents, to assess the dose that should be used for *in vivo* experiments. In the latter case, a dilution series of the chosen contrast agent (such as the small molecule dyes detailed in the materials section above) should be prepared in appropriate solvent and encapsulated into polyethylene tubes, using a glue gun to seal the tubes. The dye-filled tubes can be exchanged as required between the same phantom. Phantom imaging then proceeds as follows:

1. Place the phantom within the field of view of the OT system;
    a. Placement is best achieved using a bespoke phantom holder that will allow repeatable sample positioning between measurements.
2. Allow the phantom temperature to stabilize with the surroundings, in our experience for at least 15 minutes (longer if removing from the refrigerator). If required, apply bubble free ultrasound gel or water to couple the phantom to the ultrasound transducers.
3. Identify the regions of interest using an appropriate prescan or ‘live view’ mode. Fix the position, or if required, set translation stages to scan across multiple positions.

4. Start data acquisition.

In vivo imaging

The impact of anesthesia on the recorded signals from oxy- and deoxy-hemoglobin is significant in the organs that we tested in this study.

1. Bring animal cage to preparation area for the optoacoustic imaging system.
2. Induce anesthesia in a warm induction chamber using inhaled isoflurane at up to 3% mixed with 100% oxygen gas. Reduce isoflurane to <2% for maintenance.
3. Apply eye hydration ointment to avoid dry eyes and if in the field of view of the laser, gently apply surgical tape over the eyes.
4. For studies involving contrast agents, prepare venous catheters with identical lengths for all replicates and retain the position of the catheter inside the tail vein using suitable tissue glue.
5. Apply chosen coupling medium to the mouse (bubble free water or a thin layer of ultrasound gel) to enable good acoustic coupling.
6. Place the mouse into the OT system.
    a. *Important, use a fixed preparation time from induction of anesthesia to placement into the system. In our case this was ~ 15 min.*
7. Allow the mouse body temperature (assessed by rectal probe) to stabilize in the physiologically normal range (36-38 °C).
    a. *Important, temperature can impact respiratory rate and hence influence recorded oxy- and deoxy-hemoglobin values.*
    b. Adjust the isoflurane concentration during the recovery period (~12 min in our OT system) to maintain a respiration rate of 60-70 bpm (~1.75±0.25% isoflurane in this study).
    c. During this equilibriation period, additional setup tasks can be performed:

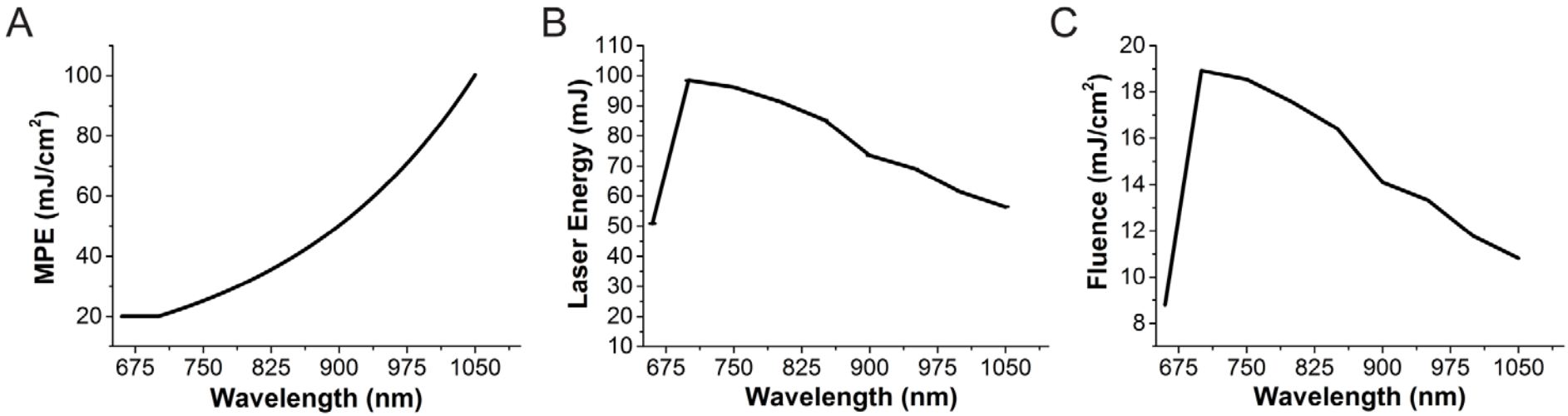
i. Check for artifacts (e.g. streaks) due to air bubbles and if found, repeat coupling and placement procedure.
ii. Identify the regions of interest to scan and if required, set translation stages to scan across multiple positions.
iii. For longitudinal imaging of a given organ, noting the position of the organ extrema and setting a central slice position can be helpful to achieve consistent anatomical positions between each replacement.

8. Start the data acquisition.
9. For reproducibility studies relating to contrast agent administration, maintain identical volumes of administration and similar perfusion time for each injection.
10. Remove the mouse at study completion, or before the imaging time duration approved by local and national animal welfare bodies is exceeded, and allow to recover in a heated recovery box.

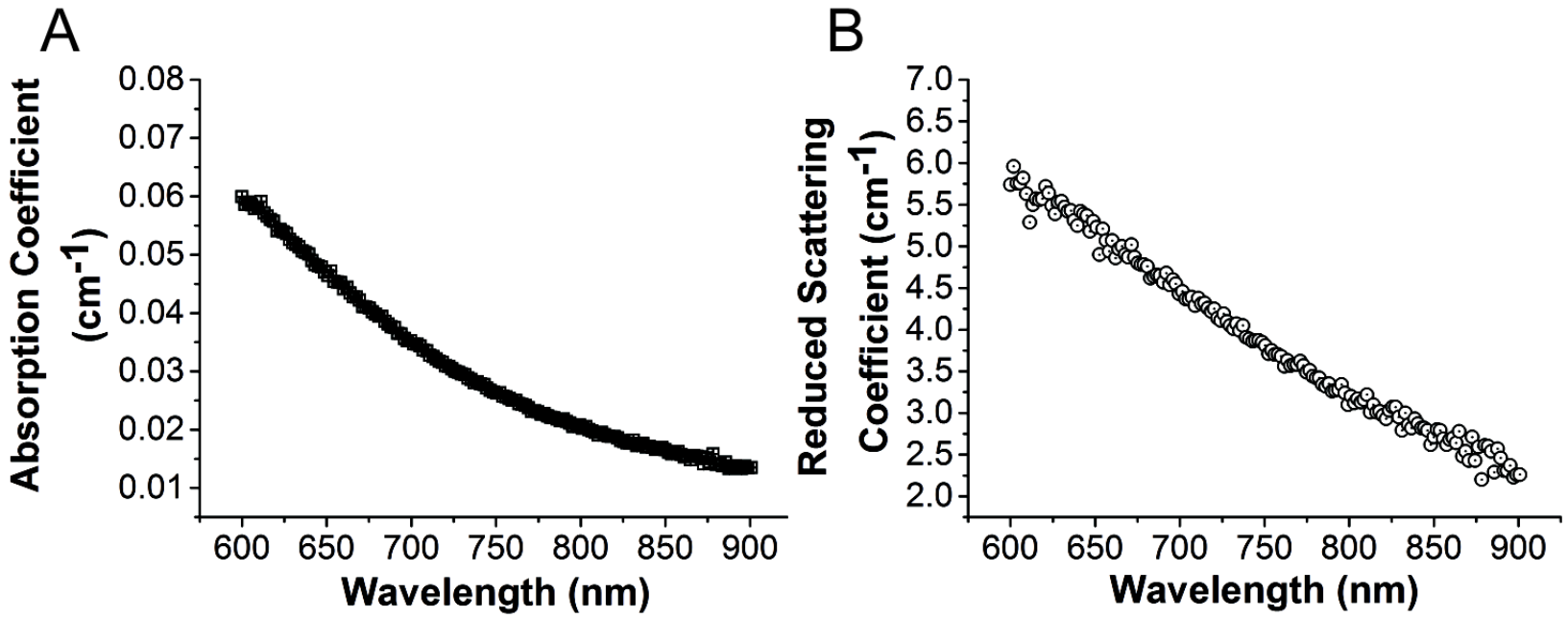
Offline image reconstruction and data analysis

Optoacoustic signals should first be corrected for the measured illumination energy to account for any wavelength-dependent variation in excitation energy. Where available, spatial and/or electrical impulse response functions for the ultrasound transducers may also be applied prior to reconstruction to account for imperfections in the ultrasound detection. All image reconstruction and analysis variables should be identified and their impact systematically evaluated for the OT system under test using an image of a stable phantom containing variable feature sizes. Where relevant, the reconstruction procedure for a given study should be applied with a consistent set of parameters, which may include: algorithm choice (e.g. backprojection or model-based); digital filters (and associated frequency settings); speed of sound in the sample; and scan conversion

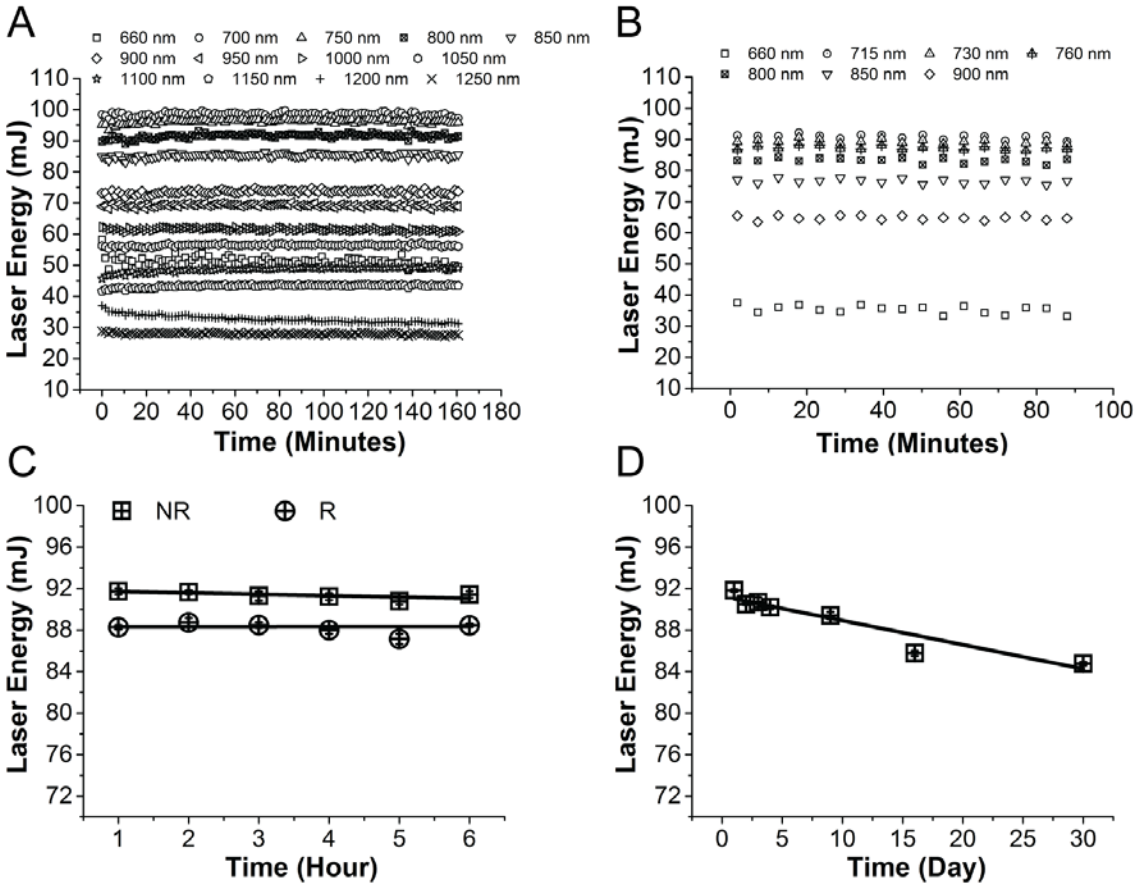
settings. The speed of sound has a significant impact on the focus of the image, so if the value for a given sample is unknown, it may need to be determined empirically in the reconstruction process. To extract quantitative data, consistent ROIs with a predefined size and location should be applied to the image and ROI statistics should be exported.



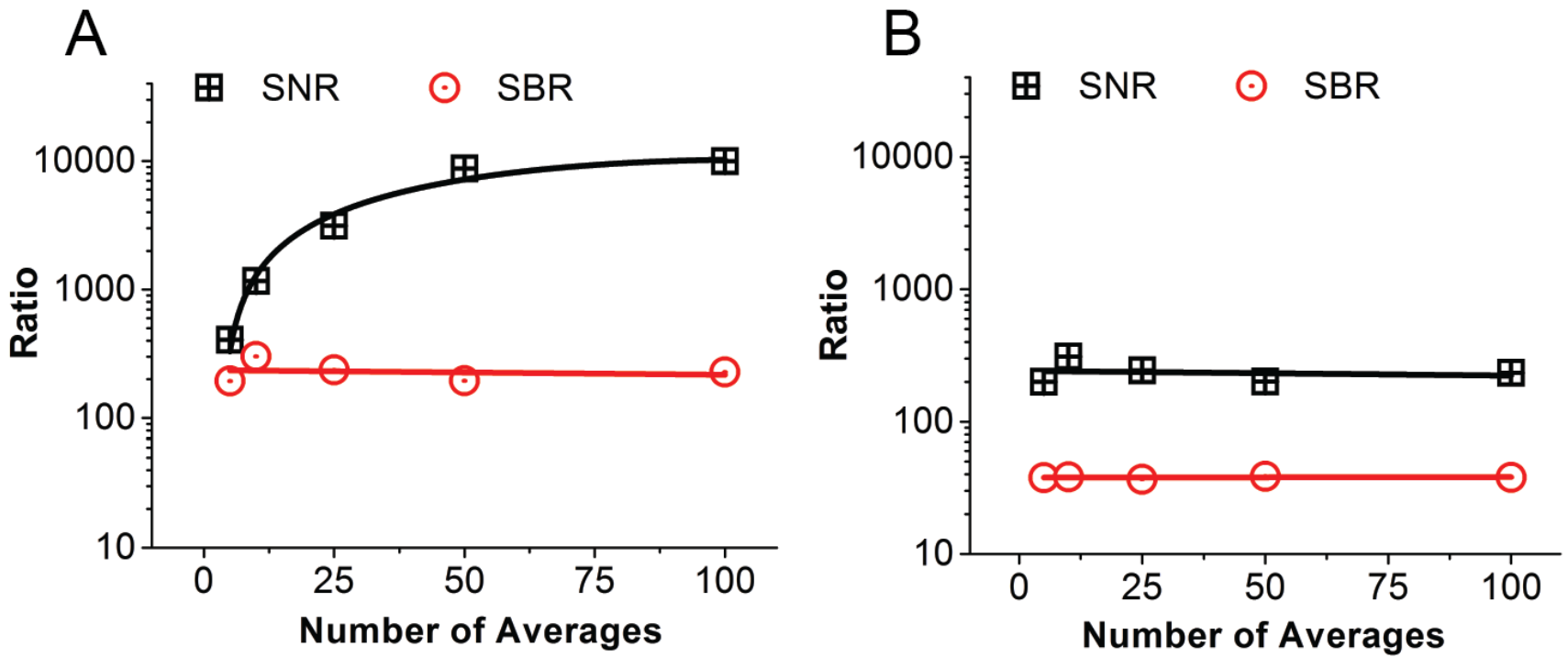
Supplemental Figure 1. Maximal permissible exposure (MPE) for skin and laser energy output over the wavelength range of interest. (A) MPE calculated according to BS EN 60825-1:1994 with IEC 60825-1 amendment 2 2001-01. (B) Laser energy measured before coupling light into the delivery fibers. (C) Corresponding light fluence at the sample, assuming the irradiated area to be the surface of a cylinder with height 8 mm (measured width of illumination) and radius 4 mm, given a beam diameter at the output of the OPO of 8.2 mm.



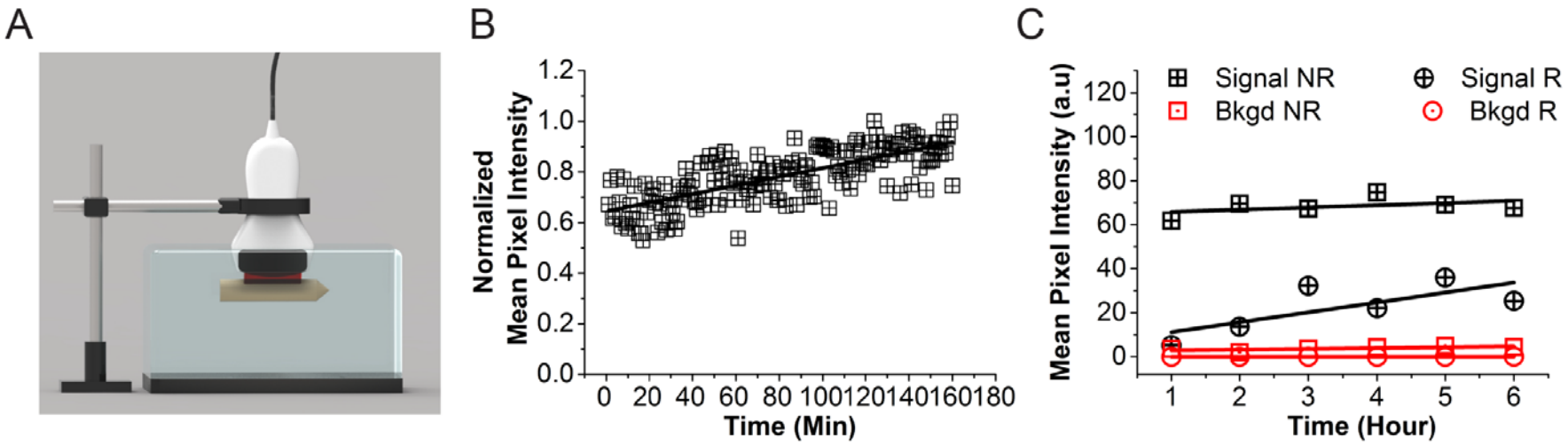
Supplemental Figure 2. Measured absorption and reduced scattering coefficients for the agarose gel composing tissue mimicking phantoms. The absorption (A) and reduced scattering (B) coefficients in the phantoms were selected based on the generic tissue optical absorption and scattering coefficients were provided for the purpose of testing devices or for protocol design by Jacques (1). The lowest suggested reduced scattering coefficient from Jacques (1), Mie $5cm^{-1}$, was chosen to facilitate direct comparison between the limits of detection measured in phantoms and *in vivo*. Data acquired using a double integrating sphere system, implemented according to published methods and calibrated against spectrophotometer measurements (2, 3). Error bars are within the points.



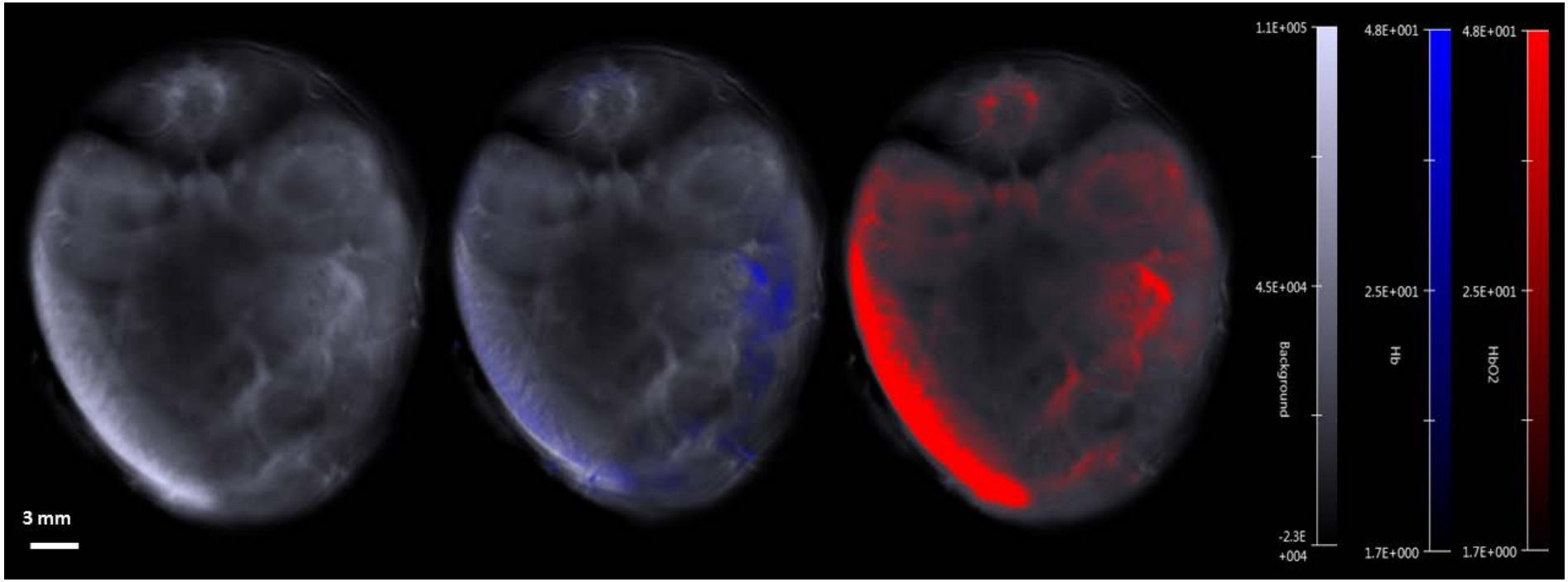
Supplemental Figure 3. Laser energy for repeatability studies. Laser energy (measured before coupling light into the delivery fibers per Figure S1B) for each wavelength used in the phantom studies shown in Figure 2 of the main manuscript over 160 minutes (A) and in the small animal repeatability study shown in Figure 4(A) of the main manuscript over 90 minutes (B). Laser energy at 700 nm for (C) 6 hour stability studies (Figure 2B) and (D) 30 day stability studies (Figure 2C); linear fits give slopes of (-13.35±6.37)E-2 mJ hour^{-1} for no replacement, (0.43±7.29)E-2 mJ hour^{-1} for replacement for (C) and (-23.27±3.15)E-2 mJ day^{-1} for (D).



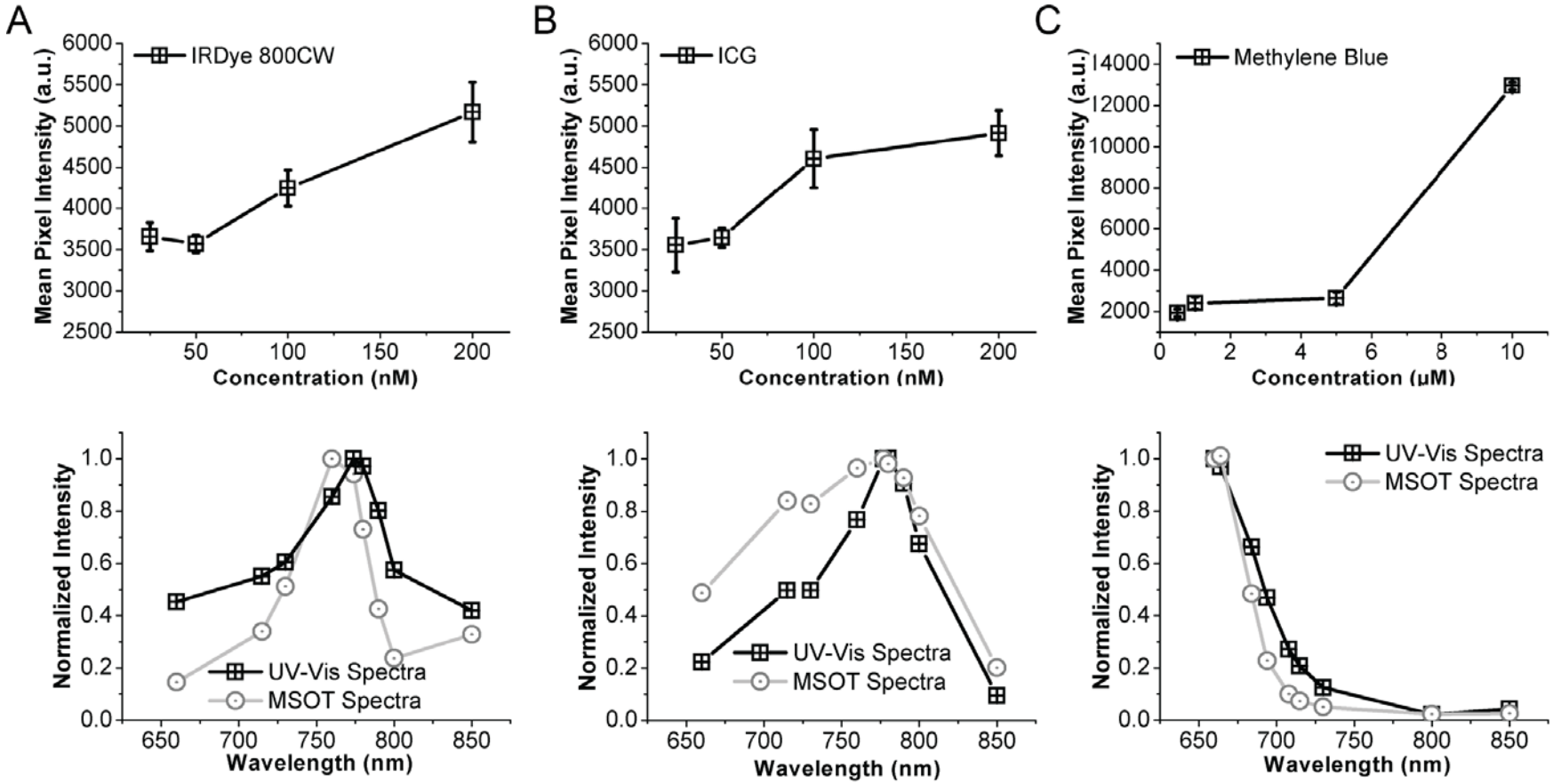
Supplemental Figure 4. Impact of frame averaging on image quality. Imaging signal-to-noise ratio (SNR) and signal-to-background ratio (SBR) as a function of the number of frame averages performed in sequential mode (A) or continuous mode (B). We calculated both parameters since we expect SNR to increase with averaging while SBR should remain constant, hence providing us with a reference. Continuous acoustic signal averaging does not influence image SNR or SBR. Polynomial fit of SNR in (A) gives B1 value of (44.35E-3±9.18E-18) and B2 value of (-2.97E-4±8.51E-20) and linear fit for SBR in (A) gives a slope of (-3.6E-4±11.4E-4) for sequential mode. Linear fits of continuous mode averaging in (B) give slopes of (-3.61E-4±1.15E-3) for SNR and (4.4E-5±1.24E-4) for SBR. Error bars are within the symbols.



Supplemental Figure 5. Optoacoustic precision as a function of time in the stable polyurethane phantom at 850 nm measured with an LED-based optoacoustic imaging system. (A) Schematic of the optoacoustic imaging system, reported in (4), and the experimental setup for phantom measurements. Briefly, optoacoustic signals are excited by a 70 ns pulsed light emitting diode array at 850 nm with a 4 kHz repetition rate and detected by a 128-element linear array transducer (9 MHz center frequency). For each excitation pulse, optoacoustic images were reconstructed with 320 signal averages. The standard operating procedure described in the Supplemental Note was applied to the same commercial polyurethane phantom used in the rest of this study, using a water bath for ultrasound coupling. (B) Normalized mean pixel intensity over 160mins exhibits a slope of (0.64±0.01) a.u. and COV of 13.9%. The upward drift may occur due to heating of the light emitting diode arrays and water medium, as the system does not currently apply per pulse energy compensation. (C) Mean pixel intensity (arbitrary units, a.u.) over 6h in a single day with replacement (R) and without replacement (NR) of the phantom between data acquisitions; slopes are (6.67±8.18) a.u. and (64.81±3.88) a.u. respectively. The COV was 6.1% over 6h (NR) rising to 51.3% (R) when the phantom was removed between data acquisitions, due to the difficulty of achieving repeatable phantom positioning.



Supplemental Figure 6. Example images of Hb (blue) and HbO_2 (red) signals. Color scales are in arbitrary units and represent the hemoglobin weighted signals derived from spectral unmixing (see Section 2.3).



Supplemental Figure 7. Sensitivity and spectral response for three commonly used dyes in tissue mimicking phantoms. Detection limits at ~1cm depth in a tissue mimicking phantom plotted as the mean pixel intensity at the peak absorption wavelength, along with absorption spectra of the NIR dyes IRDye 800 CW (A); ICG (B); and Methylene Blue (C). IRDye 800CW (IR800, Licor) and indocyanine green (ICG) were tested in concentrations ranging from 1 nM to 1 µM. Methylene blue (MB) was tested in the range from 1 nM to 50 µM. Concentration ranges relevant to the detection limit are plotted for clarity. The limit of detection was assessed by comparing the data from n=5 replicates at each dilution (from solvent only at 0nM then 1nM upwards) using a t-test until the p value for comparison of two successive dilutions was less than 0.05. The excitation wavelengths used were chosen to accurately sample

the expected excitation spectrum of the dyes according to the manufacturer published profiles: IR800: 660, 715, 730, 760, 774, 780, 790, 800 and 850 nm; ICG: 660, 715, 730, 760, 777, 780, 790, 800 and 850 nm; MB: 660, 664, 684, 694, 708, 715, 730, 800 and 850 nm. Dye characteristics were verified independently using a UV-Vis plate reader (PHERAstar FS, BMG Labtek GmbH) and the measured spectra are shown for reference. Data represent an average of n=5 MSOT scan positions per time point and error bars are within the symbols.

Supplemental Table 1: Coefficient of variation of MSOT signal intensity for the polyurethane stable phantom. Data is shown as function of excitation wavelength over 1 day, 1 week and 1 month and also for different rotational positions (see Figure 3). Larger coefficients of variation at longer wavelengths are due to increased light absorption in the water bath used for acoustic coupling.

	Coefficient of Variation (%)									
λ (nm):	660	700	750	800	850	900	950	1000	1050	1100
1 day (n=6) (with replacement)	1.2	1.2	1.3	1.3	1.4	1.3	1.8	2.8	1.7	0.9
1 day (n=6) (no replacement)	0.6	0.5	0.7	0.5	0.6	1.0	0.8	1.0	1.1	0.7
1 week (n=4)	2.0	1.9	2.0	1.9	2.0	1.9	1.7	1.6	1.6	2.2
1 month (n=7)	1.9	1.9	2.1	1.9	2.0	2.4	1.6	1.6	2.2	2.1
Positioning (n=4)	3.6	3.6	4.0	4.3	4.3	4.0	8.1	10.1	4.0	4.6

References

1. Jacques S (2013) Optical properties of biological tissues: a review. *Phys Med Biol* 58:R37-R61

2. Pickering JW, Prahl SA, van Wieringen N, Beek JF, Sterenborg HJCM and Gemert MJC (1993) Double-integrating-sphere system for measuring the optical properties of tissue. *Applied Optics* 31(4):399-410.

3. Prahl SA (2011) www.omlc.org/software/iad (Accessed 27/9/2015).

4. Toshitaka A, Naoto S, Hitoshi N, Kazuo K, Takamitsu H, Koji M, Yusuke S and Chizuyo T (2016) High frame rate photoacoustic imaging using multiple wave-length LED array light source. *Proc. SPIE* 97084E.