

<https://doi.org/10.1038/s44303-024-00061-z>

# A multiresolution approach with method-informed statistical analysis for quantifying lymphatic pumping dynamics



Mohammad S. Razavi<sup>1,2,5</sup>, Katarina J. Ruscic<sup>1,3,5</sup>, Elizabeth G. Korn<sup>3</sup>, Marla Marquez<sup>1,3</sup>, Timothy T. Houle<sup>3</sup>, Dhruv Singhal<sup>4</sup>, Lance L. Munn<sup>1</sup> ✉ & Timothy P. Padera<sup>1</sup> ✉

Despite significant strides in lymphatic system imaging, the timely diagnosis of lymphatic disorders remains elusive. This is driven by the absence of standardized, non-invasive, reliable, quantitative methods for real-time functional analysis of lymphatic contractility with adequate spatial and temporal resolution. Here, we address this unmet need by integrating near-infrared fluorescence lymphangiography imaging with an innovative analytical workflow that combines data acquisition, signal processing, and statistical analysis to integrate traditional peak-and-valley analysis with advanced wavelet time-frequency analyses. Variance component analysis was used to evaluate the drivers of variance attributable to each experimental variable for each lymphangiography measurement type. Generalizability studies were used to assess the reliability of measured parameters and how reliability improves as the number of repeat measurements per subject increases. This allowed us to determine the minimum number of repeat measurements needed per subject for acceptable measurement reliability. This approach not only offers detailed insights into lymphatic pumping behaviors across species, sex and age, but also significantly boosts the reliability of these measurements by incorporating multiple regions of interest and evaluating the lymphatic system under various gravitational loads. For example, the reliability of the peak-and-valley analysis of human lymphatic vessels was increased 3-fold using the described approach. By addressing the critical need for improved imaging and quantification methods, our study offers a new standard approach for the imaging and analysis of lymphatic function that can improve our understanding, diagnosis, and treatment of lymphatic diseases. The results highlight the importance of comprehensive data acquisition strategies to fully capture the dynamic behavior of the lymphatic system.

The lymphatic system is vital for the maintenance of tissue fluid balance and immune integrity in the human body<sup>1</sup>. Despite this, our understanding of lymphatic function has been neglected relative to other organ systems. Lymphedema is a lymphatic disease marked by abnormal accumulation of fluid in tissues. Lymphedema affects around 7 million Americans and leads to infections, chronic pain, loss of limb function, and lymphangiosarcoma<sup>2,3</sup>. Monitoring of lymphatic function and treatment response in patients with lymphatic diseases is hindered by a reliance on using disease signs for

diagnosis. At present, lymphatic diseases such as lymphedema cannot be diagnosed until the disease has progressed enough to develop early clinical signs, such as increased limb volume. While several methods such as lymphoscintigraphy, magnetic resonance imaging (MRI) lymphangiography, X-ray lymphography, fluorescein microlymphangiography and near infrared fluorescence (NIRF) lymphangiography are available at specialized medical centers to evaluate lymphatic anatomy and rudimentary lymphatic function, there are no robust, standardized techniques for imaging real-time

<sup>1</sup>Edwin L. Steele Laboratories, Department of Radiation Oncology, Massachusetts General Hospital Cancer Center, Massachusetts General Hospital and Harvard Medical School, Boston, MA, 02114, USA. <sup>2</sup>College of Engineering, University of Nebraska-Lincoln, Nebraska Lincoln, 68588, USA. <sup>3</sup>Massachusetts General Hospital Department of Anesthesia, Critical Care and Pain Medicine and Harvard Medical School, Boston, MA, 02114, USA. <sup>4</sup>Beth Israel Deaconess Medical Center, Division of Plastic and Reconstructive Surgery, Department of Surgery, Boston, MA, 02215, USA. <sup>5</sup>These authors contributed equally: Mohammad S. Razavi, Katarina J. Ruscic. ✉e-mail: [lmunn@mgh.harvard.edu](mailto:lmunn@mgh.harvard.edu); [tpadera@mgh.harvard.edu](mailto:tpadera@mgh.harvard.edu)

lymphatic contractility used in current standard medical practice<sup>4</sup>. While NIRF lymphangiography is the newest of these methods and has the most promise for measuring lymphatic contractility, there is both a lack of standardization among different camera systems<sup>5</sup>, and a lack of standardized analysis methods to optimize the functional data derived from the video data obtained with this technique. The lack of diagnostic capability and visualization of the lymphatic system prevents the timely diagnosis of lymphatic disease and impedes research for the development of novel treatments and routine monitoring of lymphatic function. Thus, there is a pressing need for improved lymphatic imaging and quantification methods to enable prompt lymphedema diagnosis and treatment, which has prompted ARPA-H to make overcoming this clinical unmet need a priority.

Collecting lymphatic vessels are comprised of tightly interconnected lymphatic endothelial cells (LECs) and lymphatic muscle cells (LMCs). The contractile unit of lymphatic vessels—the lymphangion—is the region between intraluminal valves that pumps as a result of spontaneous contractions of LMCs<sup>6,7</sup>. The unidirectional valves counteract lymph backflow when there is an opposing pressure gradient, which is affected by body positioning relative to gravity<sup>8,9</sup>. Unlike the cardiovascular system, in which intravenously introduced contrast travels to the heart, and is circulated through the entire cardiovascular tree, the unidirectional valves and distal-to-central flow of the lymphatic system mean that a central injection does not backfill and illuminate the lymphatic network in normal conditions, though dermal backflow can be observed in abnormal states<sup>10</sup>. Lymphatic imaging modalities have evolved to take advantage of interstitial or intranodal injection of tracers, which then travel to local collecting lymphatic vessels<sup>11</sup>, but do not highlight the entire lymphatic vasculature. Measurement and quantification of lymphatic contraction remain challenging, prompting ongoing evolution of innovative methods to characterize this fundamental, physiological vascular system.

Traditional lymphoscintigraphy, while considered the gold standard for lymphedema diagnosis, involves patient exposure to ionizing radiation, requires specialized imaging facilities and personnel, and affords only limited spatial resolution of the local lymphatic network<sup>12–14</sup>. This method offers some insight into lymphatic function, as decreased uptake or characteristic distributions of tracer at the endpoint of the study can occur in advanced disease. However, early lymphedema is not readily detected<sup>15</sup>. The approach does not rely on real-time measurement of lymphatic pumping, but rather pre- and post-snapshots of lymphatic network filling to calculate a lymph transit time, from which a large decrease in function can be inferred<sup>4</sup>. While X-ray lymphography and magnetic resonance lymphography afford much higher spatial anatomic resolution, like lymphoscintigraphy, they do not provide information about real-time lymphatic contractile function. Fluorescence imaging with visible light wavelengths *in vivo* is limited by signal-to-noise attenuation from tissue scattering, absorption, and autofluorescence, which prevents imaging of structures more than a few millimeters deep. In preclinical models, this depth limitation can be somewhat bypassed by removing the skin overlying the area being studied<sup>16</sup>, which is not translatable to a human bedside diagnostic tool.

NIRF lymphangiography—which utilizes excitation light greater than a wavelength of 750 nm—does not require skin removal and allows for greater depth of tissue imaging (3–4 cm) with less autofluorescence compared to fluorescence imaging with visible wavelengths<sup>17,18</sup>. Studies from 2005 to 2008 first reported its use in humans and describe the injection of indocyanine green (ICG) intradermally, which spreads locally to collecting lymphatic vessels in an anatomic region, such as a single limb<sup>19–22</sup>. While the use of ICG-based NIRF lymphangiography has expanded, it remains used primarily for anatomic mapping in humans<sup>23</sup>. NIRF technology and analysis has also been developed and investigated in preclinical models<sup>24</sup>, primarily in mice and rats. Lymphatic contractility has been quantified via NIRF lymphangiography by plotting signal intensity in a small region of interest (ROI) over time. These studies have been used for many wide-ranging, exciting applications, including detecting differences in lymphatic contraction rates between wild-type mice and those bearing B16 lymph node metastases<sup>25</sup>, untreated lymphatics in rats and those exposed to topical nitric

oxide<sup>26</sup>, normal and high salt diet in mice and rats<sup>27</sup>, and swine getting manual lymphatic drainage compared to baseline<sup>28</sup>.

A major challenge in the application of these techniques is choosing the region of interest (ROI) for analyzing the contractions. The assignment of ROIs remains a subjective manual process and there is no standardized method concerning the placement of ROIs for analyses, making it potentially difficult to compare results from different studies. Pulsatile lymph flow and the movement of “packets” of NIRF dye in the lymphatic vessels lead to increasing and decreasing fluorescent intensity over the time of a contraction. The resulting time series data can be analyzed by identifying local minima and maxima (termed peaks and valleys). The evaluation of the frequency and amplitude of signal minima is the current state-of-the-art in the lymphatic field for extrapolating information on lymphatic contraction frequency and strength in preclinical models<sup>25,26,29</sup>.

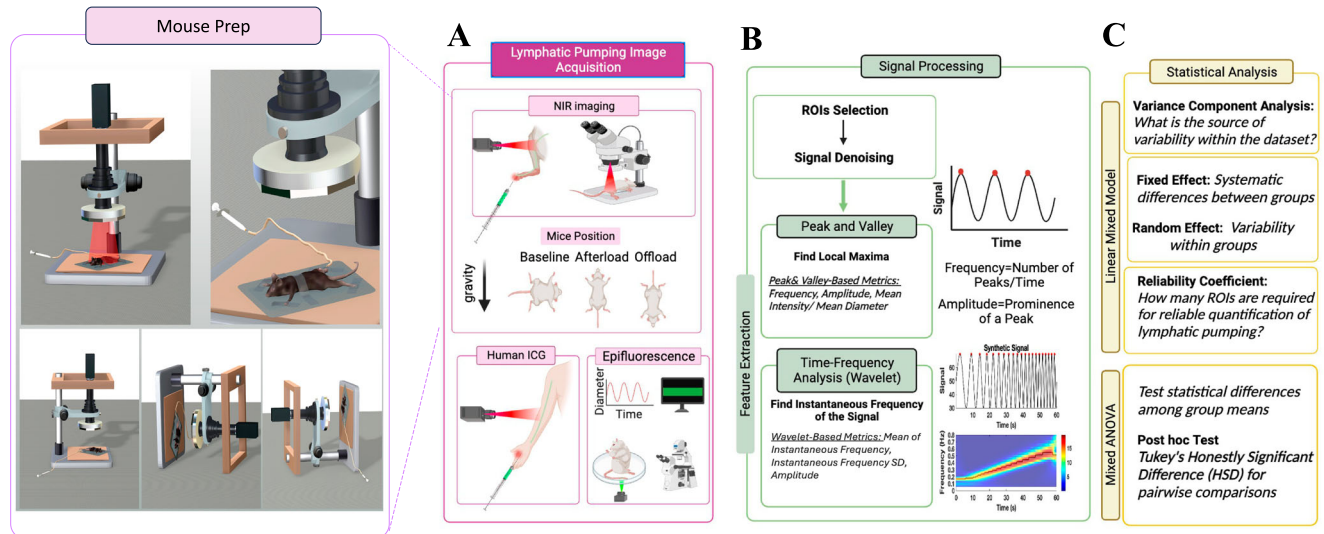
Lymphatic research has taken steps towards quantifying human NIRF lymphangiography, and several groups have applied a similar approach to human NIRF lymphangiography and quantification, but with mixed repeatability<sup>30–32</sup>. For this reason, there is a pressing need for ongoing improvement in analysis methods for the quantification of lymphatic contractility measured by NIRF lymphangiography in pre-clinical models. This would enable translation to human studies using similar methods.

In this study, we designed a workflow and developed quantitative tools to capture the different components that describe lymphatic function measured by NIRF lymphangiography. In our workflow, we consider multiple ROIs, exposing mice to various positions and loading conditions to unveil differences in lymphatic pumping behavior in both adult and aged mice of different sexes. We analyze signals using traditional peak-and-valley analysis, as well as multi-resolution wavelet-based techniques, to evaluate lymphatic pumping. The use of several ROIs is crucial for being able to make multiple measurements simultaneously from a single subject and increases the reliability of the measured lymphatic contractility. We also designed linear mixed models to account for systematic and intergroup variabilities. Our technique suggests that exposing lymphatic vessels to different loading conditions, along with using multiple ROIs, better reveals nuances in lymphatic pumping function. Our image processing and statistical methods were also applied to human NIRF data. This work lays the foundation for methods needed for future studies in humans to establish normal ranges of the parameters we study, which could then be applied in the future for disease diagnosis. This work establishes foundational methods for future human studies to more accurately define normal clinical parameter ranges for lymphatic function. These same methods could potentially be applied to disease diagnosis in the future.

## Results

### Workflow for data acquisition, signal processing, and statistical analysis for improved quantification of lymphatic pumping

One limitation of current methods, including the peak-and-valley method, is that irregular contraction patterns are not distinguished from regular contraction rhythms of the same time-averaged frequency. This limitation can introduce variability in ultimate frequencies measured when only short recordings are analyzed and also ignores what are likely important physiological states of lymphatic contractility. Our study proposes that an essential concept to make *in vivo* lymphangiography a technique that can give meaningful, reliable results, with the ultimate goal of human translation, is to collect repeated measurements from the same subject, to measure instantaneous frequencies and other contractility parameters, and to study the system in different states of behavior. Dynamic physiological systems need to be studied in a range of states to capture their essential properties. We chose gravitational stress as the impulse response system to explore the range of behavior of the lymphatic vessel network. In previously collected human ICG lymphangiography and murine intravital epifluorescence lymphangiography, data from different gravitational stresses was not available, but the power of repeated measurements and time-frequency analysis were also shown using our methods.



**Fig. 1 | Image acquisition, signal processing, and statistical analysis of real-time in vivo lymphatic pumping.** **A** In mice, after footpad injection with NIRF dye, NIRF lymphangiography is performed with the mouse and microscopy setup in three prone positions: baseline, afterload, and offload. In humans, after ICG dye injection in the interdigit webbed space, a NIRF handheld device is used to excite and image lymphatic vessels with the arm in a neutral position. In murine epifluorescence microscopy lymphangiography, FITC in the hindlimb with skin removed is imaged in a single upright position. **B** Signals are processed in MATLAB by systematically selecting a row of ROIs, denoising, and creating a peak-and-valley plot. From the plot, classical amplitude and frequency are calculated. Time-frequency wavelet

analysis is applied to the peak and valley plot, generating wavelet-based mean instantaneous frequency, instantaneous frequency standard deviation, and amplitude. **C** For each type of experiment, a variance component analysis is performed. Based on this, experiment-informed fixed effects and random effects in a linear mixed model are determined, which is then used to determine statistical differences among group means with Tukey's Honestly Significant Difference for posthoc pairwise comparisons. Decision theory is used to calculate a reliability coefficient for each metric in each experiment type to determine how many measurements per mouse or human are required for a reliable measurement.

For mouse NIRF lymphangiography gravitational studies, a custom-built frame around the NIRF imaging apparatus allowed for complete positioning of the mouse in different orientations while keeping the distance between the lymphatic vessels and excitation source and emission collection instrumentation exactly the same (Fig. 1A). For all data streams, movies were processed in MATLAB using custom code. ROIs were selected and denoised. In each ROI, the signal intensity was plotted with respect to time, creating a “peak-and-valley plot.” Local minima and maxima were identified, and frequency was calculated by the simple division of cycles (or valleys) per unit time, while amplitude was the total height of the signal from a maximum to a minimum. Note that this also relies on a user-selected threshold for amplitude minima for the peaks to distinguish a true contraction from the noise in the data in the calculation of frequency. This is the traditional method by which intravital NIRF has been analyzed, often with a single or few subjectively “best” ROI(s) chosen by the experimenter for analysis. Further, the technique is sensitive to the user-selected threshold assigned for amplitude minima (Figure S1).

To advance these methods, each ROI underwent wavelet analysis to determine the instantaneous contraction frequency over time (Fig. 1B). Mean instantaneous frequency, instantaneous frequency standard deviation, and amplitude are wavelet-based metrics that can be obtained from this analysis. Although the standard deviation of frequencies has been reported for data generated by peak-and-valley analysis, it is fundamentally a different parameter from the instantaneous frequency standard deviation described here. The standard deviations of frequency for the peak-and-valley analysis are derived from comparing multiple 300 s recordings, each yielding a single frequency. The wavelet analysis, which calculates the instantaneous frequency, gives a continuum of frequencies across the time of the recording. Therefore, there is a standard deviation for the instantaneous frequencies in a single recording. The greater the standard deviation in the wavelet analysis, the more “on” and “off” behavior is present, which is not captured by the peak-and-valley method.

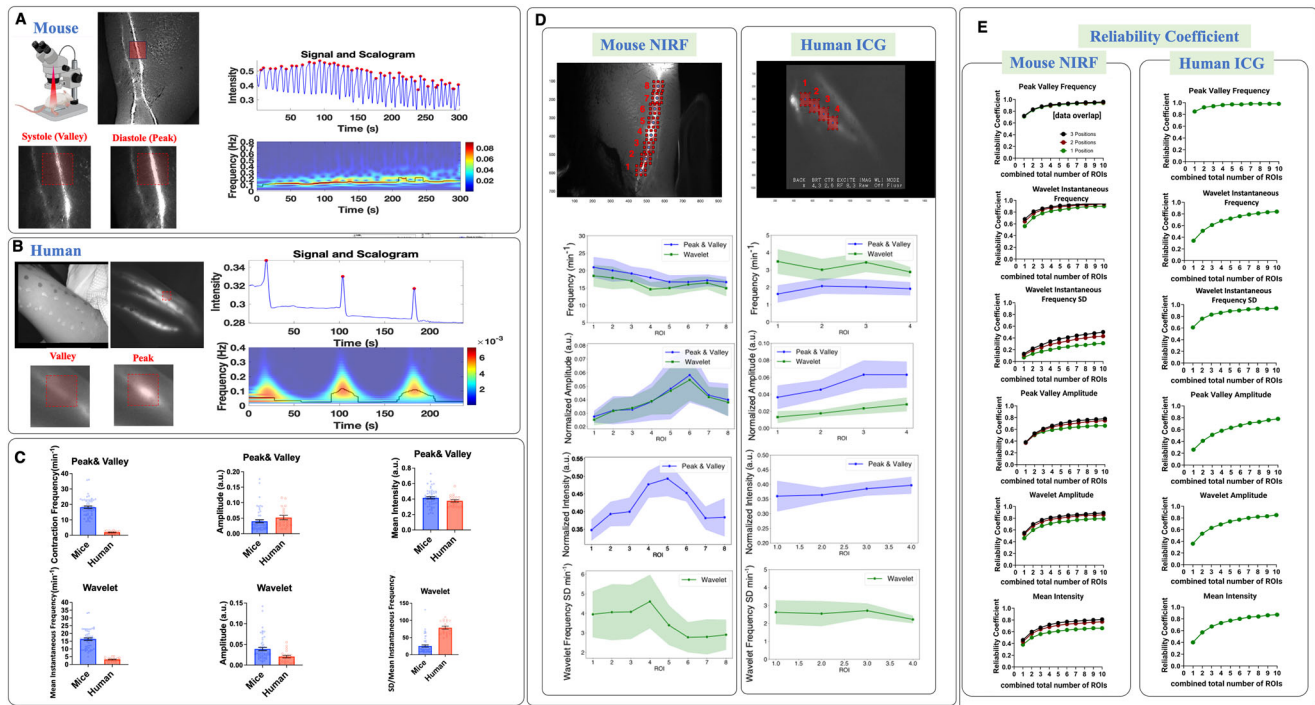
For mouse NIRF and human ICG lymphangiography, we next used statistical methods to determine the significance and reliability of our

measurements (Fig. 1C). For each method and measurement type, variance component analysis was performed to determine the sources of variability within the data (Supplementary Tables 1 and 2). Specifically, our statistical analysis revealed that the variability within individual mouse subjects accounts for 9–56% of the total variability in the measured parameters of lymphatic function. For almost all measurements, mouse-to-mouse differences are the primary source of variability in the data. This was larger than the variability attributable to age, sex, ROI, position, or interaction terms between these parameters (Supplementary Table 1). Exploration of these interactions helped determine the linear mixed model (see “Methods”; “Statistical analysis” and Supplementary Table 1). Further, the variability within human subjects accounts for 25–84% of the total variance of our human data metrics.

### Traditional peak-and-valley and wavelet time-frequency analyses reveal irregular patterns of lymphatic pumping in both mice and humans, while measurement reliability improves with the inclusion of more ROIs

NIRF imaging data was obtained from both mice and humans (Fig. 2A, B). In signal processing, ROIs were sequentially tiled along the length of the lymphatic vessel. Due to differences in tissue scattering and equipment resolution, 8 ROIs were used in recordings from mice and 4 were used in human recordings. Peak-and-valley plots of the imaging data reveal that fluorescence from the adult mouse has a continuous waveform that is both fairly regular and continuous, with a mean frequency of  $16.31 \pm 0.81 \text{ min}^{-1}$  (Fig. 2A–C). The wavelet analysis, which provides the instantaneous frequency of the signal, likewise shows a consistent instantaneous frequency with a mean of  $18.18 \pm 0.78 \text{ min}^{-1}$  for mice. On the other hand, the human peak-and-valley plot, while sometimes regular (Fig. 2B), reveals frequencies that are an order of magnitude slower than the mouse ( $3.21 \pm 0.26 \text{ min}^{-1}$  peak and valley and  $1.91 \pm 0.2 \text{ min}^{-1}$  wavelet-based mean of signal frequency). Unlike the continuous waveform of the mouse, the human waveform shows distinct peaks followed by a silent inter-peak interval (see peak/valley representative example, Fig. 2B). This is reflected in the





**Fig. 2 | Near infrared fluorescence lymphangiography measurement and analysis in mice and humans.** **A** An example region of interest (red square) in the popliteal lymphatic network in mice during systole and diastole. Peak and valley plot of the ROI showing the change in signal amplitude over time. Spectrogram of the peak and valley plot shows the instantaneous frequency over time. **B** An example region of interest (red square) in the forearm of a human patient showing the valley and the peak of ICG signal amplitude, with a corresponding peak-and-valley plot and spectrogram at that ROI. **C** The contraction frequency (from peak and valley analysis), mean intensity (from peak and valley analysis), mean instantaneous frequency (from wavelet analysis), standard deviation of instantaneous frequency (from wavelet analysis), and amplitude (from peak and valley and wavelet analysis) for mouse (left column with blue dots) and human (right column with red dots)

lymphangiography (mean  $\pm$  SEM,  $N = 8$  mice, 64 measurements,  $N = 5$  humans, 20 measurements). Mouse data are plotted only for baseline position for comparison with human baseline position. **D** Measured parameters for mouse (left) and human (right data), for peak-and-valley (blue) and wavelet (green) methods showing moderate variation in some parameters by ROI. **E** The reliability coefficient for mean intensity, wavelet frequency, wavelet instantaneous frequency SD, wavelet amplitude, peak and valley frequency, and peak and valley amplitude for mouse (left) and human (right) by number of ROIs measured for 1 (green line), 2 (red line), and 3 (black line) gravitational positions. As the values are derived from the generalizability study described in the methods, we have extrapolated reliability up to 10 ROIs, though 8 and 4 ROIs were originally measured from mice and humans, respectively.

instantaneous frequency plot from the wavelet analysis, indicated by a region of non-zero frequency (Fig. 2B) followed by a near-zero frequency between spikes.

While the frequency obtained by peak and valley versus wavelet analysis for the human data is one order of magnitude slower than for the mouse, and the peak and valley frequencies are consistent with wavelet analysis frequencies within each species (peak and valley  $18 \pm 0.78 \text{ min}^{-1}$  and wavelet  $16 \pm 0.81 \text{ min}^{-1}$  for mouse; peak and valley  $1.9 \pm 0.2 \text{ min}^{-1}$  and wavelet  $3.2 \pm 0.26 \text{ min}^{-1}$  for human), the wavelet instantaneous frequency standard deviation metric reveals much higher variations, capturing the irregular stop-start nature of the human lymphatic vessel contractility compared to mouse [instantaneous frequency normalized SD (instantaneous frequency SD  $\text{min}^{-1}$ /mean frequency  $\text{min}^{-1} \times 100$ ) human  $78.6 \pm 4.3$  vs mouse  $25.3 \pm 3.2$ ] (Fig. 2C). Additional parameters captured in the peak-and-valley analysis include the mean intensity of the ROI, which is the normalized average intensity within the ROI ( $0.42 \pm 0.01$  a.u. for mice vs.  $0.38 \pm 0.02$  a.u. for humans). The peak-and-valley amplitude, defined as peak prominence, is  $0.04 \pm 0.01$  a.u. for mice and  $0.05 \pm 0.01$  a.u. for humans. Similarly, the wavelet amplitude, quantifying the strength of the instantaneous signal, is  $0.04 \pm 0.004$  a.u. for mice and  $0.02 \pm 0.003$  a.u. for humans.

While Fig. 2A, B illustrates the behavior of a single ROI, the lymphatic vessel is a chain of lymphangions and multiple ROIs can be tiled along the chain for analysis. There is variability in lymphatic contraction frequency, amplitude, signal intensity, and instantaneous frequency standard deviation between different ROIs in both mice and humans (Fig. 2D), indicating that

different ROIs may capture different contractile behavior and that the subjective choice of a single ROI along a chain of lymphangions may skew results or fail to capture part of the functional behavior.

Using a lymphatic tissue phantom, we confirmed that our NIRF imaging system did not cause the observed differences between ROIs (Supplementary Figure S1). We also tested whether mouse skin covering the phantom contributes to the instantaneous frequency. Variations in tissue thickness, subcutaneous fat, and intradermal hair follicles affect signal amplitude, but they did not introduce any contaminant frequency signals.

To investigate how many ROIs are required for reliable quantification of pumping metrics, we used decision theory to calculate the reliability coefficient for each type of measurement (Fig. 2E). Increasing the number of measurements per mouse, both by increasing the number of ROIs and by increasing the number of gravitationally different positions, increases the reliability of the measurement (Fig. 2E). For example, wavelet frequency calculated using only one ROI in one position has a reliability coefficient of only 0.56 and this increases up to 0.95 with eight ROIs and three body positions. This analysis also helped us identify at which point increasing the number of measurements per mouse has diminishing returns. For example, for peak and valley frequency, increasing the number of ROIs in a single position from one to five increases reliability from 0.71 to 0.91, while increasing the number of ROIs from five to 10 only increases the coefficient further to 0.94, a gain that may not be experimentally significant at the cost of increasing analysis time.

Each measurement type must be individually considered. For example, the measurement for wavelet instantaneous frequency standard deviation has high variability and even with many ROIs and three positions, the

reliability coefficient curve in mice does not saturate, predicting that additional measurements per mouse beyond those performed in this study would continue to increase the reliability of wavelet instantaneous frequency standard deviation. For our human data, only one position was measured. Compared to the analogous measurements in mice, certain metrics, such as wavelet frequency and peak-and-valley amplitude, show a lower reliability coefficient in the human data for low numbers of ROIs. However, it is important to note that the reliability coefficients for peak and valley frequency and wavelet frequency SD are slightly higher in the human data. For example, the mouse reliability coefficient for wavelet frequency with four ROIs is 0.82, while the human reliability coefficient is 0.68. As seen in the murine data, increasing the number of ROIs increases reliability, with some metrics reaching saturation, while more variable measurements do not reach saturation even with 10 modeled ROIs. This suggests that the number of ROI measurements per human subject must be maximized to help improve the reliability of measurements and may be greater than the number needed in mice.

An important consideration is how to analyze repeat measurements in subjects when performing hypothesis testing between groups. One approach that has been used in the past is simply to use the mean of all individual data measurements of a single type when comparing groups, or to create a single data point for each measurement type in a subject by averaging the repeat measurements for that subject. In both approaches, valuable information is lost. Here, we employed a statistical approach that uses linear mixed models, in which repeat measurements in subjects are preserved as unique numbers and subject grouping can be achieved by using random intercepts by-subject in the model (see “Methods”; “Statistical analysis”).

### Changes in position induces contractility variations between mice across different ages and sexes compared to baseline position

We next performed NIRF imaging and analyses with mice in different positions with respect to gravity. For each video from a mouse in a single position undergoing NIRF imaging, eight ROIs were manually selected in MATLAB and tiled systematically in a row along the lymphatic vessel (Fig. 3A). Measurements were then generated using both peak-and-valley and wavelet analysis. We analyzed the datasets several times with slight variations in the position of the eight tiled ROIs to test whether user input for ROI impacts results, and the dataset did not substantively change with slight variations in the tiling (data not shown). Linear mixed effects models including ROI, sex, age, position, and some interaction terms between these (see “Methods”; “Statistical analysis”) were used with Tukey’s Honestly Significant Difference (HSD) test for pairwise comparisons. In addition to the baseline position (mouse flat and prone; Figs. 1A and 3A), we measured the mouse contractility with the head up (maximal afterload on the lymphatic vessel) and head down (maximal offload on the lymphatic vessel) (Figs. 1A, 3A). In the afterload position, we hypothesized that the hindlimb lymphatic vessel would experience the greatest load as the vessel must pump against closed valves with the maximal column of fluid pressure above. In the offload position, the valves point in the same direction as the gravitational force. We hypothesized that the measured lymphatic vessel would experience the lowest load in this configuration. Surprisingly in the adult mouse (Fig. 3B), there was no difference in contraction frequency between the afterload position ( $20.3 \pm 3.8 \text{ min}^{-1}$  for peak-and-valley and  $20.0 \pm 3.2 \text{ min}^{-1}$  for wavelet) and baseline measurements ( $19.2 \pm 3.8 \text{ min}^{-1}$  for peak-and-valley and  $17.3 \pm 3.2 \text{ min}^{-1}$  for wavelet). Compared to baseline prone position, the amplitude of pumping and mean intensity of the signal in the afterload position is lower in adult mice (Fig. 3B), consistent with the vessel being able to move less fluorescent dye in these increased load conditions. We did not see a recovery of amplitude or signal intensity when mice were returned from the afterload to the offload position, though the frequency did decrease.

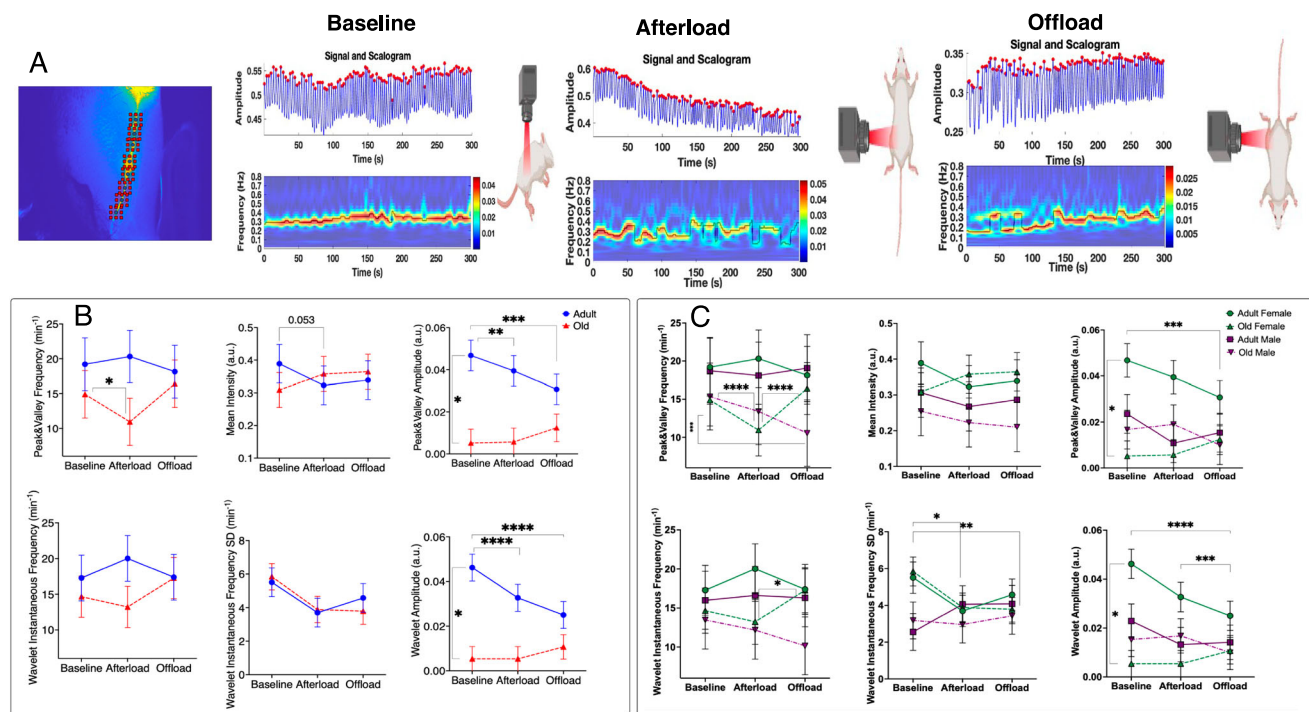
The behavior of old mice is different than adult mice (Fig. 3B), with overall frequencies and amplitudes lower in all three positions for old mice.

While baseline position measurements are different between age groups, the stress of the afterload position reveals an even larger difference between the groups. When faced with afterload, the contraction frequency decreases in old mice (old mouse afterload to old mouse baseline contrast for peak-and-valley frequency  $\text{min}^{-1}$ : mean difference =  $-2.95$ ,  $\text{SE} = 0.77$ ,  $p = 0.002$ ). This demonstrates the value of measuring the behavior of the dynamic physiological system of lymphatic vessels under multiple load conditions to explore the full gamut of system behavior. The decrease in lymphatic contraction frequency in old mice under afterload conditions could be interpreted as pump failure due to the additional load (i.e., the system is unable to compensate for the increased load by increasing pumping strength). In Fig. 3B, in both peak-and-valley and wavelet frequency graphs, the old mice (red line) show a pattern of decreased pumping frequency with afterload position compared to baseline position, with a recovery increase in the offload position (not all of these comparisons reach significance in difference in Fig. 3B). We further analyzed our data by sex, which revealed additional differences between sexes (Fig. 3C), including that the lymphatic pumping behavior splits by sex in the old mice. Notably, the recovery of peak-and-valley frequency in the offload position in old mice is only in female mice (green dotted line) ( $p < 0.001$ ). When male and female mice are analyzed together, the recovery of frequency in offload is not significant. The old male mice (dotted purple line in Fig. 3C) continue to decrease in frequency in the offload position and there is a significant decrease in peak-and-valley frequency between old male mice in the baseline and offload positions (peak and valley frequency  $\text{min}^{-1}$  baseline to offload: mean difference =  $4.79$ ,  $\text{SE} = 1.2$ ,  $p = 0.0053$ ).

### Wavelet analysis of murine intravital epifluorescence microscopy reveals differences in lymphatic pumping regularity not revealed by traditional analysis methods

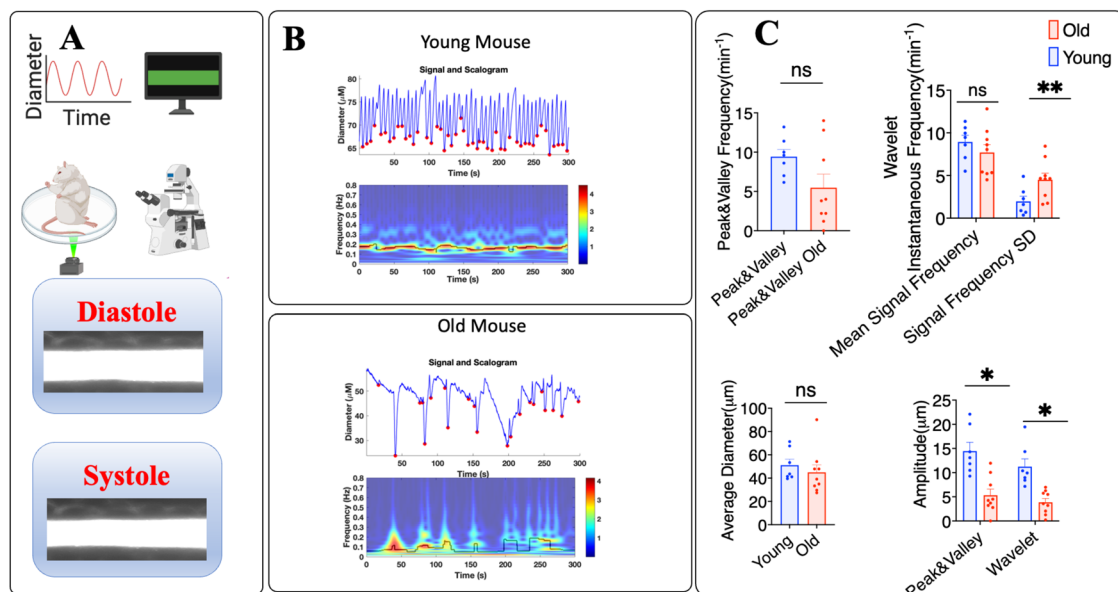
To test whether our wavelet analysis can be applied to lymphatic contractility measurements from intravital epifluorescence microscopy, we re-analyzed previously published intravital epifluorescence microscopy data in young and old mice<sup>33</sup> that was previously quantified using only peak-and-valley analysis (Fig. 4A). Both traditional peak-and-valley and wavelet analyses suggest there is no statistically significant difference in frequency between young and old mice. For young mice, the value was  $8.3 \pm 0.9 \text{ min}^{-1}$ , and for aged mice, it was  $7.7 \pm 0.9 \text{ min}^{-1}$ , based on wavelet analysis. Similarly, using peak-and-valley analysis, the values were  $8.5 \pm 1.2 \text{ min}^{-1}$  for young mice and  $5.5 \pm 1.7 \text{ min}^{-1}$  for aged mice. The value of frequency obtained by peak and valley analysis is also dependent on a user-selected threshold of peak change, in this re-analysis  $2 \mu\text{m}$  of vessel diameter was used, but choosing a larger threshold would have likely led to a statistical significance based on age. We have provided a sensitivity analysis demonstrating this dependency in Supplementary Figure S2. Both peak and valley techniques and wavelet analysis show that the amplitude of contractions significantly decreases with aging ( $13.9 \pm 1.7 \mu\text{m}$  in young and  $4.5 \pm 1.1 \mu\text{m}$  in old for peak-and-valley; and  $10.4 \pm 1.6 \mu\text{m}$  in young and  $3.6 \pm 0.8 \mu\text{m}$  in old for wavelet,  $p < 0.05$ , unpaired t-test).

While the peak-and-valley plot visually looks irregular in the old mice, the peak and valley plot is smooth and regular in young mice (Fig. 4B). Previous methods for quantifying lymphatic contractility fail to capture this change to more irregular contraction behavior with age. Interestingly, the wavelet instantaneous frequency standard deviation, which measures the deviation from the mean of the instantaneous frequency, is significantly higher in the old mouse group due to large changes in the instantaneous frequency of lymphatic pumping ( $1.9 \pm 0.5 \text{ min}^{-1}$  for young vs  $4.8 \pm 1.0 \text{ min}^{-1}$  old group,  $p < 0.01$ ) (Fig. 4C). This novel parameter can capture irregularities such as the start-stop nature of lymphatic contraction in aged mice. In this re-analysis of prior data, only one ROI per mouse was available, so we did not explore reliability analysis with multiple repeat measures. In intravital epifluorescence microscopy, time-frequency analysis using wavelet analysis is a powerful tool that can be especially useful in the analysis of aged or diseased mice to capture changes in the regularity of lymphatic pumping.



**Fig. 3 | Near infrared fluorescence lymphangiography measurement in mice across age, sex, and lymphatic position.** **A** An example image from a series of pictures comprising the 5-min lymphangiography recording showing 8 regions of interest tiled across the lymphatic vessel. The peak-and-valley plot and spectrogram are shown in the baseline position, afterload position (head up), and offload position (head down). **B** The model estimate  $\pm$  SEM from linear mixed modeling for peak and valley frequency ( $\text{min}^{-1}$ ), wavelet frequency ( $\text{min}^{-1}$ ), peak and valley amplitude (a.u.), wavelet amplitude (a.u.) and peak and valley mean intensity (a.u.) are given for

the baseline, afterload and offload positions in adult (blue line) and old (red line) mice ( $N = 8$  adult mice,  $N = 8$  old mice), data collapsed across sex. Tukey's HSD was used to determine statistical differences between group means for posthoc pairwise comparisons. **C** The same data as in **(B)** is shown split by sex, with green solid line adult female ( $N = 5$ ), green dotted line old female ( $N = 4$ ), purple solid line adult male ( $N = 3$ ), purple dotted line old male ( $N = 4$ ). Tukey's HSD was used to determine statistical differences between group means for posthoc pairwise comparisons.



**Fig. 4 | Intravital epifluorescence microscopy.** **A** A lymphatic vessel afferent to the popliteal lymph node in a live mouse is filled with fluorescein isothiocyanate dye by interdermal injection of the mouse foot and imaged on an epifluorescence microscope. A single region of interest is captured and diameter changes over time are recorded. An example lymphatic vessel in diastole and systole is shown. **B** Example data in a young mouse. The average vessel diameter over time is measured over 300 s. The spectrogram below shows the instantaneous frequency throughout the

recording in an adult mouse. Below, example of diameter vs. time and instantaneous frequency vs. time for an old mouse. **C** Mean  $\pm$  SEM of contraction frequency (peak-and-valley analysis), frequency (peak-and-valley analysis), amplitude (peak-and-valley and wavelet analysis), and average diameter (peak-and-valley analysis) for young mice (blue) ( $n = 9$ ) and old mice (red) ( $n = 8$ ). Unpaired t-test was used to determine statistical differences between group means for posthoc pairwise comparisons.



## Discussion

The past two decades have witnessed advancements in non-invasive imaging modalities for imaging lymphatic vessels. Among these techniques, NIRF lymphatic imaging and ICG lymphography, which are closely related techniques, have gained popularity for applications in small animal models of lymphatic disease and the clinical evaluation of human lymphatic dysfunction, respectively<sup>17,22,24,25,28,34</sup>. Typically, the analysis of NIRF signals is performed by selecting an ROI and analyzing the average signal intensity over time, which is associated with the intrinsic contraction of the lymphatic vessels. Functional metrics based on peak-and-valley analysis, such as frequency, ejection fraction, emptying rate, pumping pressure and fractional pump flow, have been used to analyze NIRF signals in lymphatic diseases<sup>35,36</sup>. While these metrics offer valuable insights, the peak-and-valley-based metrics lack temporal resolution.

Some recent studies show the benefit of more advanced signal processing techniques for understanding lymphatic contractility. For example, wavelet analysis has shown promise for demonstrating the entrainment of lymphatic contraction in an isolated vessel experimental setup with applied cyclic fluid pressure<sup>37</sup>. Wavelet analysis is commonly used for electroencephalograms (EEGs)<sup>38</sup>, and has also been used to analyze and extract features of complex biomedical signals such as electrocardiograms (ECGs)<sup>39,40</sup> and electromyograms (EMGs)<sup>41,42</sup>. Recently, wavelet analysis has been applied to analyze lymphatic dysfunction in a lymphedema mouse model<sup>43</sup>, but this work focused on the dominant frequency, which was also identified by finite Fourier transforms. Unlike Fourier transforms, which are suitable for temporally stationary signals, wavelet transforms—with their varying frequency and length—are better suited for signals with time-varying components, thus providing multi-resolution characteristics that provide insights into both temporal and frequency aspects of the signal. In particular, wavelet analysis can be useful when there is irregularity in lymphatic contractility. The novel contractility parameter of instantaneous frequency SD can capture the irregularity of lymphatic contractility, which is missed in traditional peak-and-valley analysis. In addition, while the peak-and-valley analysis depends on a user-defined threshold for minimal diameter changes to detect peaks (Figure S1), wavelet analysis eliminates this need. In scenarios where weaker contraction amplitudes are anticipated, such as in aged subjects, frequency domain techniques like wavelet analysis could offer the advantage of eliminating potential user bias and reducing variability across different experimental datasets. While we have demonstrated that wavelet analysis can be applied to both NIRF and intravital epifluorescence lymphangiography signals in mice, it is important to recognize that some differences in measurements are expected between these methods, even though both aim to capture intrinsic lymphatic properties. Intravital epifluorescence lymphangiography requires surgical removal of the overlying skin, which likely introduces multiple physiological changes not present in the non-invasive NIRF method, such as inflammation, tissue edema, and alterations in neurohormonal regulation.

In NIRF imaging studies of the human lymphatic system, the parameters of lymphatic propulsion frequency and velocity are valuable and innovative physiological metrics that have been used for assessing lymph flow<sup>44,45</sup>. However, the calculation of these parameters necessitates the measurement of the distance traveled by fluorescent markers. The data captured in NIRF imaging is in the form of 2-dimensional images, but the fluorescence originates from a 3-dimensional lymphatic vessel that varies in position along the Z-plane of the human tissue. Consequently, the 2-dimensional distance derived from projecting photons from all Z-planes in the collected data overlooks changes in the tissue depth of the imaged vessel. This omission restricts the reliability of velocity comparisons between individuals due to the unaccounted variance in vessel depth that leads to different angles of the vessel relative to the imaging plane. Therefore, to ensure a more consistent and comparative analysis across different subjects, we prioritize the frequency and amplitude of fluorescence as our principal derived parameters, as they offer a more straightforward comparison between subjects without the complexities introduced by depth variations of lymphatic vessels. Further, imaging vessels that reside deeper in tissue will

have more light scattering effects that can add to measurement variability between subjects.

While most studies have used a single or few ROIs when analyzing lymphatic contractility via NIRF imaging, our statistical analysis using a linear mixed model suggests that such quantification may not be reliable. Reliability coefficient testing assigns a value between 0 to 1 to a given test, with  $R = 1.00$  indicating perfect reliability<sup>46</sup>. Test-retest reliability in animal research has great variability, with certain measurements notoriously unreliable between different experimenters and groups, such as in behavioral studies<sup>47,48</sup>. For clinical measurements used in patient care, a reliability coefficient above 0.9 has been proposed as acceptable<sup>49</sup>. To achieve reliable measurements for most of our experimental parameters in mice, we need at least five measurements from different ROIs, while more variable metrics require an even greater number of ROIs. The measurement reliability is further increased by studying lymphatic vessels in multiple positions with respect to the force of gravity.

In the published literature, there is significant variability in the reported lymphatic contractility in mice, which is a function of the measurement technique (Figs. 3 and 4), anesthesia<sup>50</sup>, body position<sup>51</sup>, volume of contrast<sup>51</sup>, mouse strain<sup>52</sup>, the specific lymphatic vessel (e.g., mesenteric vs. IALV) and, likely, other parameters that we do not yet understand. Resolving the contribution of any individual input parameter becomes more challenging if the analysis itself adds variability. One goal of developing the more robust analysis tools for the measurement of lymphatic contractility was to reduce a source of experimental variability that will allow the more effective interrogation of the inputs that mediate lymphatic vessels contractility.

The use of multiple ROIs is advantageous as it allows the analysis of lymphatic pumping at different positions along the chain of lymphangions. For our human lymphangiography measurements, the highest number of measurements that were available (4 ROIs) did not meet clinical reliability standards of a reliability coefficient  $>0.9$  for our functional metrics, demonstrating that development of a reliable bedside lymphangiography test requires more measurements per subject than is typically performed for ICG lymphangiography anatomic studies. We showed that decision theory can be used to find the optimal number of measurements per subject needed to create a reliable measurement for each experimental parameter, and that this number differs between mice and humans, even when using similar NIRF lymphangiography techniques. It is notable that in both mice and humans, lymphatic contractility greatly varies between subjects, and subject variability is a dominant factor in the total attributable measurement variability of our experiments. This highlights the need for making measurements within a single subject as reliable as possible. We showed that increasing the number of ROIs measured and handling the repeated measurements thoughtfully with a mixed model is a way to increase reliability. Increasing reliability is key in the development of tests that can be safely implemented in clinical decision-making. We also found that variance component analysis can be used to identify which parameters in the data (for example, subject number) contribute to the greatest source of variability in the data. Variance component analysis then informs an optimal linear mixed model for statistical analysis between different groups.

Our data also show differences in lymphatic contractility parameters between mice and humans. The overall frequency of lymphatic contraction in the human data is lower than that of adult mice, while the irregularity of contractility as measured by normalized frequency SD is greater. There is emerging evidence that anesthetics change lymphatic contractility<sup>50</sup>, and notably, mice were under anesthesia during lymphangiography while humans were not. While the large differences in lymphatic function between mice and humans are most likely driven by species differences, anesthetic factors are worth bearing in mind and testing in greater detail in the future.

The lymphatic system needs to adjust to different fluid loading scenarios. Gravity is one important factor that could alter fluid dynamics in the interstitial space and vessels, creating different hydrostatic pressure gradients (e.g., in standing versus supine positions). In clinical settings, it is common to elevate the lower part of a patient's body above their head in some surgical procedures, which exposes lower extremity lymphatic vessels

to favorable hydrostatic pressure gradients and head-and-neck lymphatics to unfavorable pressure gradients. Experiments simulating the effects of microgravity on lymphatic contractility, along with computational modeling, also suggest that gravity alters lymphatic contractility<sup>53,54</sup>. In the present study, we utilized NIRF imaging and designed a specialized microscope fixture to investigate the effects of gravity on lymphatic contractility. Employing a quantitative approach, we assessed the differences in lymphatic function between older and younger mice when placed in various positions relative to gravity. Our analysis indicates that exposing adult mice to a head-up position, which maximizes the effect of gravity on the lower limbs, had limited effect on lymphatic contractility. Similarly, placing adult mice head-down, where gravity facilitates flow by inducing a favorable hydrostatic pressure, does not significantly alter contraction frequency. This is likely due to the small size of mice, which limits the magnitude of the gravitational forces. However, older mice exhibit decreased contractility in the head-up position compared to their baseline, revealing age-related differences in response to gravitational changes. We suggest incorporating measurements in different gravitational positions to standard NIRF mouse lymphangiography to more fully capture the range of behavior of lymphatic contractility. Furthermore, splitting the data by individual sex shows differences in some metrics between groups, suggesting that sex is an important biological variable in lymphatic contractility quantification.

In lymphatic measurements, a persistent challenge and objective is to minimize animal-to-animal variability<sup>55</sup> and handle regional heterogeneity<sup>56</sup>. NIRF lymphangiography in mice has been criticized for its unreliability and high variability compared to ex vivo lymphatic measurement methods<sup>57</sup>. We show that incorporating more measurements per animal, in the form of more ROIs and measurements in multiple body positions can generate highly reliable data with test-retest coefficients of experimental parameters >0.9, considered adequate even for clinical tests. In the emerging field of human ICG lymphangiography, increasing measurements per human subject may be what is needed to evaluate ICG lymphangiography as a reliable clinical tool that could be utilized in clinical practice to quantify lymphatic function in at-risk patients and those already with the burden of lymphatic disease. Larger future studies in healthy human volunteers are needed to establish the normal range of parameters defined in this work by utilizing our multi-ROI approach and the statistical methods we describe here. ICG lymphangiography could also be a valuable tool for monitoring lymphedema progression and evaluating therapeutic responses—particularly with enhancement in reliability provided by our new methodology. Finally, wavelet transform analysis is a powerful method to reveal irregularity in lymphatic contraction that is not revealed by traditional lymphangiography methods, which is especially crucial in the study of human subjects and mice with disease phenotypes. In order to better integrate these techniques in experimental or clinical workflows, image acquisition, parameter extraction and multi-ROI wavelet analysis will need to be automated with modern computational tools. Combined, implementing these new methods can bring human and murine NIRF lymphangiography to the forefront as a more reliable tool in testing lymphatic function in patients and preclinical models.

## Methods

### Mouse model

C57BL/6 mice, female and male, 6 months old, were obtained from the Cox-7 animal facility operated by the Edwin L. Steele Laboratories, Department of Radiation Oncology at the Massachusetts General Hospital (MGH). C57BL/6 mice, female and male, 18 months old, were obtained from the National Institute on Aging. Both ages of mice were housed at the MGH Center for Comparative Medicine facilities, Charlestown Navy Yard. Animal protocols were approved by the Institutional Animal Care and Use Committees (IACUC) at MGH, and all facilities are accredited by the Association for Assessment and Accreditation of Laboratory Animal Care International (AAALAC).

### Imaging lymphatic pumping

A combination of ketamine and xylazine was used to anesthetize mice. The fur was removed using a depilatory cream, and 5  $\mu$ L of Dextran (10 K), Flamma® 774 (BioActs) was injected intradermally into the dorsal aspect of the mouse paw. NIRF imaging, as previously described<sup>51</sup>, was used to image the saphenous lymphatic vessels afferent to the popliteal lymph node. Briefly, the imaging system is equipped with a  $\times 6.5$  Zoom lens (Navitar), a Prosilica GT2750 camera (Allied Vision Technology), and an ICG-B emission filter (832/37, Semrock). A 760 nm laser-emitting diode (Marubeni), filtered through a 775/50 bandpass filter (Chroma), was used to excite the dye. To study how gravity affects lymphatic pumping, a custom-designed microscope fixture was integrated with the imaging system. This setup allowed the positioning of mice to experience different gravitational forces during imaging. The recordings were conducted in 5-min intervals, during which the mice were positioned in three distinct postures: (1) prone, (2) head-up (their heads elevated above their bodies), (3) head-down (their heads positioned lower than their bodies).

### Human ICG lymphangiography

De-identified videos were obtained from Dr. Dhruv Singhal at BIMDC as previously described. Briefly, 0.1 mL stock solution (2.5 mg/mL) ICG solution (Akorn) was sterilely mixed with 2.5 mg albumin and injected intradermally. Multiple injections were introduced as previously described<sup>23</sup>. An injection overlying the cephalic vein was introduced to image the upper arm. Only upper arm images were used in this study. A NIRF imager (Hamamatsu PDE Neo II, Mitaka USA) was used to visualize superficial lymphatic vessels with the ‘mapping’ mode of the device. Institutional review board approval was obtained (2021P000209) for this prospective lymphangiography analysis. Consent was obtained from every patient to perform the studies.

### Signal processing

A custom-written MATLAB code was used to analyze images recorded at 10 fps. The ROIs, consisting of a  $50 \times 50$  pixel area (mouse) or  $100 \times 100$  pixel area (human), were selected along the lymphatic vessels. The normalized average intensity ( $0 < \text{Intensity} < 1$ ) within each ROI was recorded at each time point. To denoise the signal, we used a wavelet denoising technique based on the Block James-Stein method, which yields optimal global and local adaptivity<sup>58</sup>. The denoised signal was analyzed based on two approaches: (1) peak-and-valley: frequency was detected based on the local minima in the signal, and amplitude was determined based on peak prominence, which measures how the peak’s height stands relative to other peaks. (2) Multiresolution time-frequency analysis: the instantaneous frequency of the signal was detected through the wavelet transformation of the signal, identifying the wavelet component with the highest magnitude at each time point. The instantaneous amplitude of the signal was also correlated with the magnitude of the corresponding wavelet component with the highest magnitude. Specifically, continuous wavelet transformation employing the Morse mother wavelet was utilized, which is well-suited for analyzing non-stationary signals that exhibit frequency variations over time. The average and standard deviation of instantaneous frequency as well as the amplitude in the instantaneous frequency and amplitude of the signal were used as metrics for lymphatic pumping.

## Detailed procedures

### Mouse preparation

1. Place the mouse on a scale to weigh it. Anesthetize the mouse using a ketamine-xylazine mixture (100 mg/10 mg per kg body weight) injected subcutaneously in the mouse’s lower left quadrant.
  - *Note: Maintain the body temperature at 37°C throughout the procedure using a heating pad set to 39°C.*
2. Once the mouse is in an adequate plane of anesthesia, apply ophthalmic ointment to both eyes, then shave the mouse’s right hindlimb



toward the thigh initially and then in the opposite direction, toward the footpad, for optimal fur clearance.

3. Remove any remaining hair from the right lower hindlimb by using depilatory cream. Remove the cream after 60 s with filter-sterilized saline, and wipe to prevent chemical skin burns.
4. Inject 5  $\mu$ L of Dextran (10 K) Flamma® 774 (BioActs) intradermally into the dorsal aspect of the mouse paw, between the 2nd and 3rd digits.
  - *Note: Injections on the ventral aspect of the footpad are usually subcutaneous rather than intradermal, take longer to fill the lymphatic network, and are thus not preferred.*

### Mouse staging

5. Place the mouse on a rectangular-shaped light-controlled aluminum sheet in the prone position and tape down all extremities and the tail with the right hindlimb at a 45-degree angle away from the body—place tape along the upper back to secure the mouse for positional changes.
6. Tape the light-controlled aluminum sheet onto the warming pad with dual-sided mounting tape and frame the edges with tape to ensure the sheet is pressed firmly against the heating pad.

### Imaging

7. The afferent lymphatic vessels to the popliteal lymph node are imaged using near-infrared fluorescence imaging using a previously described NIRF setup<sup>51</sup>.
8. The imaging system is composed of a  $\times 6.5$  Zoom lens (Navitar) and a Prosilica GT2750 camera (Allied Vision Technology) with an ICG-B emission filter (832/37, Semrock).
9. A custom-built ring light source is used to excite the dye using 760 nm high-power laser-emitting diodes (Marubeni) filtered through a 775/50 bandpass filter (Chroma).
10. A custom-written MATLAB code is used to analyze images. The code is available upon request from the authors.

### Positional maneuvers

11. Take a 5-min recording of the mouse in the Baseline position (imaging system flat).
12. Move the apparatus using the supportive wooden frame so the mouse is in the afterload position and allow 2 min for equilibration, then take a second 5-min recording.
13. Move the apparatus again to place the mouse in the offload position, allow 2 min for equilibration, and then take a third 5-min recording.

### Tissue phantom

To create a lymphatic tissue phantom, an 18 cm length of PE tubing (0.28 mm inner diameter) was fixed to the bottom of a 60 mm diameter petri dish with cellophane tape. Polydimethyl siloxane (PDMS) was prepared by mixing Sylgard 184 (Dow Corning) base with curing agent at a ratio of 10:1. The PDMS liquid was poured over the PE tubing in the petri dish to a thickness of 3 mm, and then cured in an oven at 60 °C for 2 h. A 32 g needle was inserted into the PE tubing for infusion of the contrast agent. The tubing was filled with the same Flamma-774 NIR dye used in the mouse experiments in the paper. The mouse hindlimb was prepared as in the NIRF experiments (shaved and treated with depilatory cream from the popliteal region to the ankle). The hairless skin was harvested, placed on top of the tissue phantom and NIRF emission from the dye-loaded tissue phantom was measured under no-flow conditions, with and without the harvested skin, to assess instrument noise and tissue scattering.

### Statistical analysis

**Variance components analysis and generalizability theory.** The primary aims of this study were to evaluate the percent of variance

attributable to each measurement facet (mouse, position, ROI, age, sex) for each outcome (wavelet frequency and instantaneous standard deviation, wavelet amplitude, mean intensity, and peak-and-valley frequency and amplitude) and to assess the reliability of the measurements. The attributable variance (expressed as a percentage) for each measurement facet was assessed for each outcome using a linear mixed effects model that included each facet as its own random intercept, as well as random intercepts for all two-way interactions between facets (Initial Model in Supplementary Table 1). Within-mouse two-way interactions that were not crossed (e.g., mouse:sex and mouse:age) were excluded as they could not be estimated because each mouse was only one sex and one age.

To assess measurement reliability, the full model from above was adjusted for each outcome to remove any within-mouse interaction terms that had zero or near-zero variance (Refined Model in Supplementary Table 1). The reliability of a measurement is determined by the proportion of the observed score variance attributable to variance among the “true” scores. Generalizability studies are designed to isolate an object of measurement and estimate its variation<sup>59</sup>. Using the mouse identifying number (Mouse\_ID) as the object of measurement, the simplified model for each outcome was input into the “gstudy” and “dstudy” functions of the “gtheory” package in RStudio. Relative error variances, an estimate of the “noise” around a measurement, were calculated for different numbers of positions and ROIs using variance components from the “gstudy” output. Relative error variance was calculated as the sum of the variance components related to position, ROI, and position  $\times$  ROI. Generalizability coefficients, which are akin to reliability coefficients, were then estimated for each ROI using these relative error variances. The generalizability coefficient was calculated as the variance component related to the facet of interest (e.g., ROI  $\times$  sex) divided by that variance component plus relative error. For a full elaboration of the methods see Huebner and Lucht<sup>60</sup>. The generalizability coefficients were used to assess reliability as a function of the number of ROIs and position measurements.

**Hypothesis testing.** The secondary aim of this study was to examine associations between mouse characteristics and the near-infrared fluorescence imaging outcomes. Descriptive statistics for continuously scaled variables are reported as mean  $\pm$  standard error of the mean. These data were analyzed using a linear mixed effects model, with a random intercept by mouse to account for the repeated nature of data collection. Models also included a two-way multiplicative interaction between ROI and sex, and a three-way interaction between age, sex, and position. The following structure in R was used:

$$\text{Outcome} \sim \text{ROI} + \text{Sex} + \text{Age} + \text{Position} + \text{ROI}*\text{Sex} + \text{Sex}*\text{Age} + \text{Age}*\text{Position} + \text{Sex}*\text{Position} + \text{Sex}*\text{Age}*\text{Position} + (1|\text{Mouse\_ID})$$

Model estimates are reported as predicted mean  $\pm$  standard error. Tukey’s HSD test was used to assess pairwise comparisons of the three-way interaction between age, sex, and position, and the two-way interaction between age and position. Results were evaluated using two-sided  $p$ -values, with values less than 0.05 considered statistically significant.

Power calculations were conducted to illustrate the sample size necessary to detect a statistically significant difference between age groups (adult vs. old) for each outcome for reliability = 0.90 and reliability = 0.70. These values were chosen to represent high and low levels of reliability in outcome measurements. Low levels of reliability obscure the differences between groups causing a reduction (i.e., attenuation) of observable effect size difference. To obtain the expected effect sizes that would be observed under these reliability conditions, the observed effect sizes and reliabilities from each model were disattenuated and attenuated to generate theoretical effect sizes. In this way, the effect sizes that would be observed under different levels of reliability could be calculated. To estimate power, a two-sample independent  $t$ -test was used to determine the  $n$  per group needed to achieve 0.80 power with a type I error rate of 0.05.

## Data availability

Data presented in this study are available on request from the authors though restrictions may apply based on permission from the Massachusetts General Hospital and/or the Beth Israel Deaconess Medical Center. Data will be available on reasonable request pending permissions and data use agreements from sending and receiving institutions.

## Code availability

The code is available upon request from the authors.

Received: 6 June 2024; Accepted: 30 November 2024;

Published online: 17 January 2025

## References

- Kesler, C. T. et al. Lymphatic vessels in health and disease. *Wiley Interdiscip. Rev. Syst. Biol. Med.* **5**, 111–24 (2013).
- Sleigh B. C. & Manna, B. *StatPearls* (StatPearls Publishing, 2023).
- Sharma, A. & Schwartz, R. A. Stewart-Treves syndrome: pathogenesis and management. *J. Am. Acad. Dermatol.* **67**, 1342–8 (2012).
- O'Donnell, T. F. et al. New diagnostic modalities in the evaluation of lymphedema. *J. Vasc. Surg. Venous Lymphat Disord.* **5**, 261–273 (2017).
- Zhu, B. et al. Determining the performance of fluorescence molecular imaging devices using traceable working standards with SI units of radiance. *IEEE Trans. Med. Imaging* **35**, 802–11 (2016).
- Padera, T. P., Meijer, E. F. & Munn, L. L. The lymphatic system in disease processes and cancer progression. *Annu. Rev. Biomed. Eng.* **18**, 125–58 (2016).
- Scallan, J. P. et al. Lymphatic pumping: mechanics, mechanisms and malfunction. *J. Physiol.* **594**, 5749–5768 (2016).
- Skalak, T. C., Schmid-Schönbein, G. W. & Zweifach, B. W. New morphological evidence for a mechanism of lymph formation in skeletal muscle. *Microvasc. Res.* **28**, 95–112 (1984).
- Schmid-Schönbein, G. W. Microlymphatics and lymph flow. *Physiol. Rev.* **70**, 987–1028 (1990).
- Rockson, S. G. Diagnosis and management of lymphatic vascular disease. *J. Am. Coll. Cardiol.* **52**, 799–806 (2008).
- Polomska, A. K. & Proulx, S. T. Imaging technology of the lymphatic system. *Adv. Drug Deliv. Rev.* **170**, 294–311 (2021).
- Sherman, A. I. & Ter-Pogossian, M. Lymph-node concentration of radioactive colloidal gold following interstitial injection. *Cancer* **6**, 1238–40 (1953).
- Celebioglu, F. et al. Lymph drainage studied by lymphoscintigraphy in the arms after sentinel node biopsy compared with axillary lymph node dissection following conservative breast cancer surgery. *Acta Radiol.* **48**, 488–495 (2007).
- Mango, L. Lymphoscintigraphic indications in the diagnosis, management and prevention of secondary lymphedema. *Radiation* **3**, 40–45 (2023).
- Mihara, M. et al. Indocyanine green (ICG) lymphography is superior to lymphoscintigraphy for diagnostic imaging of early lymphedema of the upper limbs. *PLoS ONE* **7**, e38182 (2012).
- Liao, S. et al. Method for the quantitative measurement of collecting lymphatic vessel contraction in mice. *J. Biol. Methods* **1**, e6 (2014).
- Sevick-Muraca, E. M., Kwon, S. & Rasmussen, J. C. Emerging lymphatic imaging technologies for mouse and man. *J. Clin. Invest.* **124**, 905–14 (2014).
- Sevick-Muraca, E. M. Translation of near-infrared fluorescence imaging technologies: emerging clinical applications. *Annu Rev. Med.* **63**, 217–31 (2012).
- Kitai, T. et al. Fluorescence navigation with indocyanine green for detecting sentinel lymph nodes in breast cancer. *Breast Cancer* **12**, 211–5 (2005).
- Unno, N. et al. Preliminary experience with a novel fluorescence lymphography using indocyanine green in patients with secondary lymphedema. *J. Vasc. Surg.* **45**, 1016–21 (2007).
- Sevick-Muraca, E. M. et al. Imaging of lymph flow in breast cancer patients after microdose administration of a near-infrared fluorophore: feasibility study. *Radiology* **246**, 734–41 (2008).
- Rasmussen, J. C. et al. Lymphatic imaging in humans with near-infrared fluorescence. *Curr. Opin. Biotechnol.* **20**, 74–82 (2009).
- Granoff, M. D. et al. Variable anatomy of the lateral upper arm lymphatic channel: an anatomical risk factor for breast cancer-related lymphedema. *Plast. Reconstr. Surg.* **152**, 422–429 (2023).
- Kwon, S. & Sevick-Muraca, E. M. Noninvasive quantitative imaging of lymph function in mice. *Lymphat. Res. Biol.* **5**, 219–232 (2007).
- Proulx, S. T. et al. Use of a PEG-conjugated bright near-infrared dye for functional imaging of rerouting of tumor lymphatic drainage after sentinel lymph node metastasis. *Biomaterials* **34**, 5128–37 (2013).
- Weiler, M., Kassiss, T. & Dixon, J. B. Sensitivity analysis of near-infrared functional lymphatic imaging. *J. Biomed. Opt.* **17**, 066019 (2012).
- Kwon, S. et al. Altered lymphatic function and architecture in salt-induced hypertension assessed by near-infrared fluorescence imaging. *J. Biomed. Opt.* **17**, 080504–1 (2012).
- Sharma, R. et al. New horizons for imaging lymphatic function. *Ann. N. Y Acad. Sci.* **1131**, 13–36 (2008).
- Robinson, H. A. et al. Non-invasive optical imaging of the lymphatic vasculature of a mouse. *JoVE Mar.* **8**, e4326 (2013).
- Groenlund, J. H. et al. A validation study of near-infrared fluorescence imaging of lymphatic vessels in humans. *Lymphat. Res. Biol.* **15**, 227–234 (2017).
- Granoff, M. D. et al. A novel approach to quantifying lymphatic contractility during indocyanine green lymphangiography. *Plast. Reconstr. Surg.* **144**, 1197–1201 (2019).
- Thorup, L. et al. The transport function of the human lymphatic system — a systematic review. *Physiol. Rep.* **11**, e15697 (2023).
- Lei, P. J. et al. Lymphatic muscle cells are unique cells that undergo aging induced changes. Preprint at *bioRxiv* <https://doi.org/10.1101/2023.11.18.567621> (2023).
- Proulx, S. T. et al. Quantitative measurement of lymphatic function in mice by noninvasive near-infrared imaging of a peripheral vein. *JCI Insight* **2**, e90861 (2017).
- Weiler, M. & Dixon, J. B. Differential transport function of lymphatic vessels in the rat tail model and the long-term effects of Indocyanine Green as assessed with near-infrared imaging. *Front Physiol.* **4**, 215 (2013).
- Mukherjee, A. et al. Lymphatic injury alters the contractility and mechanosensitivity of collecting lymphatics to intermittent pneumatic compression. *J. Physiol.* **599**, 2699–2721 (2021).
- Mukherjee, A. et al. Entrainment of lymphatic contraction to oscillatory flow. *Sci. Rep.* **9**, 5840 (2019).
- Coifman, R. R. *Wavelet Analysis and Signal Processing* (1990).
- Engin, M. Spectral and wavelet based assessment of congestive heart failure patients. *Comput Biol. Med.* **37**, 820–8 (2007).
- Yang, H., Bukkapatnam, S. T. & Komanduri, R. Nonlinear adaptive wavelet analysis of electrocardiogram signals. *Phys. Rev. E* **76**, 026214 (2007).
- Phinyomark, A., Limsakul, C. & Phukpattaranont, P. *Application of Wavelet Analysis in EMG Feature Extraction for Pattern Classification* (2011).
- Angkoon, P., Pornchai, P. & Chusak, L. In *EMG Methods for Evaluating Muscle and Nerve Function*, Ch. 7 (ed. Mark, S.) (IntechOpen, 2012).
- Cheon, H. et al. In vivo dynamic and static analysis of lymphatic dysfunction in lymphedema using near-infrared fluorescence indocyanine green lymphangiography. *Arterioscler Thromb. Vasc. Biol.* **43**, 2008–2022 (2023).
- Tan, I. C. et al. Assessment of lymphatic contractile function after manual lymphatic drainage using near-infrared fluorescence imaging. *Arch. Phys. Med. Rehabil.* **92**, 756–764.e1 (2011).

45. Zhang, J. et al. Automated analysis of investigational near-infrared fluorescence lymphatic imaging in humans. *Biomed. Opt. Express* **3**, 1713–23 (2012).
46. Matheson, G. J. We need to talk about reliability: making better use of test-retest studies for study design and interpretation. *PeerJ* **7**, e6918 (2019).
47. Mandillo, S. et al. Reliability, robustness, and reproducibility in mouse behavioral phenotyping: a cross-laboratory study. *Physiol. Genomics* **34**, 243–55 (2008).
48. Gouveia, K. & Hurst, J. L. Optimising reliability of mouse performance in behavioural testing: the major role of non-aversive handling. *Sci. Rep.* **7**, 44999 (2017).
49. Portney, L. G. & Watkins, M. P. *Foundations of Clinical Research: Applications to Practice* (2015).
50. Bachmann, S. B. et al. Differential effects of anaesthesia on the contractility of lymphatic vessels in vivo. *J. Physiol.* **597**, 2841–2852 (2019).
51. Bouta, E. M. et al. Lymphatic function measurements influenced by contrast agent volume and body position. *JCI Insight* **3**, e96591 (2018).
52. Liao, S. et al. Impaired lymphatic contraction associated with immunosuppression. *Proc. Natl. Acad. Sci. USA* **108**, 18784–9 (2011).
53. Gashev, A. A., Delp, M. D. & Zawieja, D. C. Inhibition of active lymph pump by simulated microgravity in rats. *Am. J. Physiol. Heart Circ. Physiol.* **290**, H2295–308 (2006).
54. Li, H. et al. Computational simulations of the effects of gravity on lymphatic transport. *PNAS Nexus* **1**, pgac237 (2022).
55. Gashev, A. A., Davis, M. J. & Zawieja, D. C. Inhibition of the active lymph pump by flow in rat mesenteric lymphatics and thoracic duct. *J. Physiol.* **540**, 1023–37 (2002).
56. Gashev, A. A. et al. Regional heterogeneity of length-tension relationships in rat lymph vessels. *Lymphat Res. Biol.* **10**, 14–9 (2012).
57. Davis, M. J. et al. Lymphatic contractile dysfunction in mouse models of Cantú Syndrome with K(ATP) channel gain-of-function. *Function* **4**, zqad017 (2023).
58. Chicken, E. Block-dependent thresholding in wavelet regression. *J. Nonparametr. Stat.* **17**, 467–491 (2005).
59. Webb, N. M. & Shavelson, R. J. Generalizability theory. In *Wiley StatsRef: Statistics Reference Online* (2014).
60. Huebner, A. & Lucht, M. Generalizability theory in R. *Pract. Assess. Res. Eval.* **24**, 5 (2019).

## Acknowledgements

This work was supported by NIH grants F32HL156654 (M.S.R.), R21AG072205 (T.P.P.), R01HL128168 (L.L.M., T.P.P.), R01CA284372 (T.P.P.), R01CA284603 (L.L.M., T.P.P.), R21EB031982 (L.L.M.), R01CA247441 (L.L.M.), K08GM155886 (K.J.R.), Department of Defense PRMRP grant (HT94252410100) and support from the Rullo Family MGH Research Scholar Award from the MGH Research Institute (T.P.P.).

This work was also supported by the National Institutes of Health via Harvard Anesthesia T32-GM007592, by a grant from the International Anesthesia Research Society awarded to K.J.R., and by the Harvard Medical School Eleanor and Miles Shore Award to K.J.R. Biorender was used in the creation of figures.

## Author contributions

M.S.R. and K.J.R.: Experimental design, data collection, data figures, data analysis, manuscript writing, and manuscript revision. E.G.K.: Statistical analysis and manuscript revision. M.M.: Data collection, data figures, and manuscript revision. T.T.H.: Statistical design, statistical analysis, and manuscript revision. D.S.: Data collection and manuscript revision. L.L.M.: Experimental design, figure design, and manuscript revision. T.P.P.: Experimental design and manuscript revision. All authors reviewed the final manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains

supplementary material available at

<https://doi.org/10.1038/s44303-024-00061-z>.

**Correspondence** and requests for materials should be addressed to Lance L. Munn or Timothy P. Padera.

**Reprints and permissions information** is available at

<http://www.nature.com/reprints>

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

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

© The Author(s) 2025