Supplementary Material

Materials and Methods:

Wound Tissue Sample Acquisition and Histological Processing

All experiments were approved and performed according to the University of Arkansas IACUC (Protocols #16001, #17063). C57BL/6J mice (Between 1-16 months of age; male & female) were given carprofen (5 mg/kg, s.c.), and 6 mm full-thickness, excisional wounds were produced on the dorsum using a sterile biopsy punch. Wounds were longitudinally monitored and mice were euthanized either at days 3, 5, or 10 post wounding. Approximately 1 cm^2 of skin wound tissue was excised from the mice following euthanasia. Excised tissue was flash frozen in Tissue-Tek® optimal cutting temperature compound (Sakura Finetek; Tokyo, Japan) at -80°C. The frozen wound tissue was sectioned in a Leica CM1860 cryostat (Wetzlar, Germany) into 30 µm thick samples, transferred to glass slides, and stored at -80°C before being regressively stained with hematoxylin and eosin (H&E). H&E stained slides were originally imaged in their entirety at 4x magnification on a Nikon Ni-Eclipse (Tokyo, Japan). A total of 31 H&E stained slides were imaged, with 25 slides being used to create the training, validation, and test image sets. The remaining 6 slides (2 per post-wounding day) were not involved in the training process and were used for evaluation of the network's performance on whole slide segmentation.

Network Design and Training Regime

From the 25 H&E stained slides, 395 unique 512x512 pixel images were isolated to encompass various regions of wounded skin tissue (between 5-25 images/slide). The image set was augmented by a simple y-axis reflection leaving a final pool of 790 images total. Each image was manually traced using custom Matlab code into masks of 7 encompassing regions of epidermis, dermis/hypodermis, granulation/wound tissue, scab, hair follicles, sub-cutaneous skeletal muscle,

and background light. The dermis and hypodermis regions were combined into one collective class as measurements of wound dimension are not inherently dependent on subcutaneous adipose tissue. Once traced, 70% of the images were assigned to the training set, 20% to the validation set, and 10% to a testing set using a random number generator (Figure S1b). The network's design followed a U-Net architecture with 4 symmetric encoding and decoding layers and is diagramed in Figure S1a. Convolutions used a 3x3 pixel element and a ReLu activator. The network was created and trained at an initial learning rate of 10^{-3} to 100 epochs using an Adam optimizer on Matlab 2019a. The number of training epochs was decided by monitoring validation loss at the end of each epoch. When validation loss ceased to decrease, training was ended to prevent overfitting. Training took approximately 16 hours on a Windows 64bit OS machine housing a GeForce RTX 2070.

Segmentation Network Assessment on Test Images and Whole Wound Sections

Once the network was trained, it was used to segment the set of 79 test images not involved in the training process. The network predictions were compared to each manually segmented, ground truth test image on a pixel-by-pixel basis. These comparisons were represented as confusion matrices to understand class-specific accuracies. Accuracy is represented along the diagonal of the confusion matrix, and off-diagonal elements provide insight into which classes cause the network some confusion. Overall total accuracy of the network was calculated as the sum of the correctly predicted pixels relative to the total number of pixels in the test set.

In addition to the test set, the network was also evaluated on its ability to segment whole tissue sections. The 6 slides not used to create training data were fully segmented using a sliding 512x512 pixel element across the entirety of the image. For each 512x512 input only the 256x256 pixel region in the center was kept to avoid any edge artifacts. A post processing step was also

applied to eliminate noise in the granulation mask by isolating and filling the binary granulation mask in areas not contested by the background, muscle, or hair follicles classes. Network accuracy was assessed using the same approach as the test set.

Automated Quantification of Wound Metrics

A custom Matlab code was created to automatically calculate wound depth, wound width, epidermal/dermal thicknesses, and percentage of re-epithelialization from user and network segmented H&E slides. Wound width was calculated as the minimum separation distance between the left and right portions of segmented dermis transformed with a discrete Laplacian (the del2 function in MATLAB). The wound width bounds were also used to find the percentage of re-epithelialization by measuring the fraction of the bounded granulation mask covered by pixels classified as epidermis. The thicknesses of the epidermis and dermis/hypodermis were calculated using a Euclidean distance transform via the bwdist function in MATLAB. Thickness measurements were both computed as an average and as a function of the long axis of the tissue section (Figure S2). Correlations between the network and ground truth thickness measurements (Figure S2) were measured using a Pearson's correlation coefficient (R). All correlation and percent error measurements are presented as an average ± the standard deviation based on all 6 tissue sections.

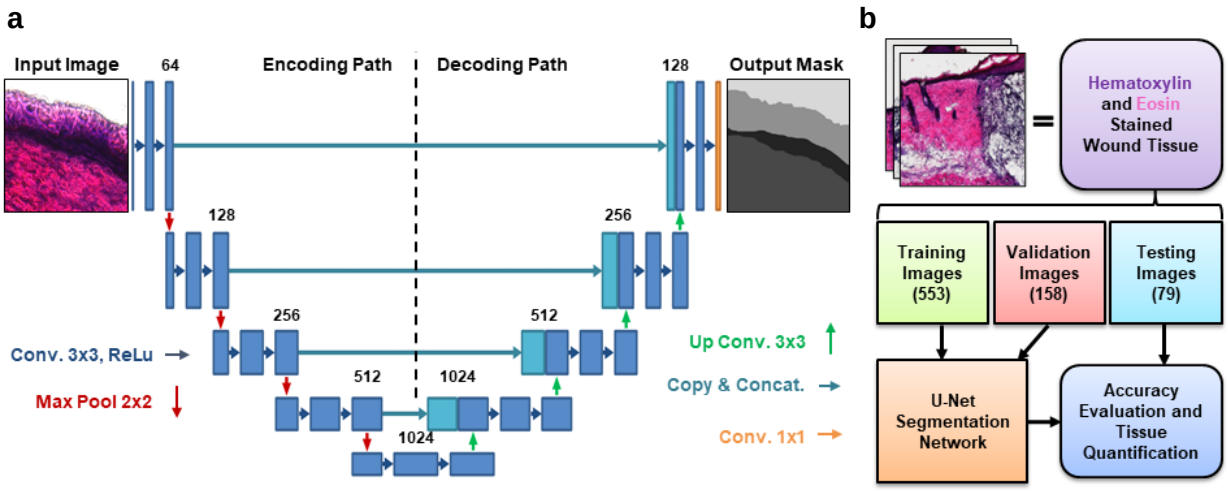

Figure S1. Outline of network architecture and training process. (a) The network was comprised of a symmetric set of 4 encoding and decoding layers following a U-Net architecture. Numerical values above each layer indicate the number of filters and subsequent feature maps. (b) A total of 790 H&E stained images acquired at 4x magnification and cropped to be 512x512 pixels (~610x610 µm) were randomly assigned to a training, validation, and testing set following a 70%, 20%, 10% distribution. The network was trained exclusively on the images allocated to the training set, while images in the validation set provided updates on validation loss during training to prevent network overfitting. Once trained, the testing image set was used to independently evaluate network accuracy.

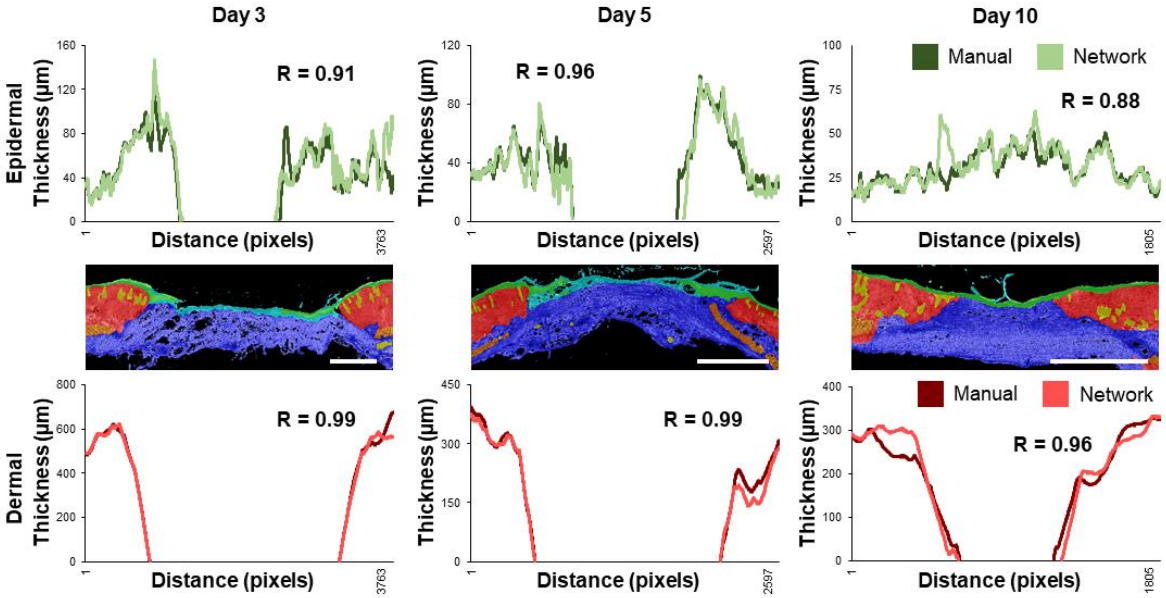

Figure S2. Epidermal and dermal thicknesses measured across the long axis of the wound sections. Epidermal and dermal thickness values were calculated along the x-axis using a Euclidean distance transform of the manually segmented (dark green/red) and network segmented (light green/red) wound sections. The Pearson's correlation coefficient (R) was found between the thickness plots and is presented as the average R for each post-wound time point. All scale bars are 500 µm.