Supplementary Methods for "Automated image analysis identifies signaling pathways regulating distinct signatures of cardiac myocyte hypertrophy", Bass et al.

Automated myocyte segmentation

The algorithm for automated myocyte segmentation is composed of 6 main phases, which are outlined in Figure 1. In the first phase (Figure 1-1), the algorithm loads images of nuclei (stained using DAPI) and α-actinin, a myocyte-specific cytoskeletal protein. A median filter with an appropriately sized window (4 pixels in the experiments presented here) is applied to the α-actinin channel to improve cell segmentation performance, and then all images are background-subtracted using a manual threshold. In the second phase (Figure 1-2), nuclei are identified and segmented using an Otsu threshold in the DAPI channel.

The Otsu threshold [10] is used to classify pixels as either objects or background using the global properties of the image histogram without operator input. Ideally, the object and background pixels would show up as two distinct peaks on the image histogram with a clear valley in between separating these two categories, making threshold selection intuitive. However, in real images the separation point between these two peaks is often muddled by noise. The Otsu threshold is the grey level in an intensity histogram that maximizes the variance between background and foreground pixels, which is also the threshold that minimizes the intra-class variance of background and foreground pixels. In the Otsu algorithm, first the image histogram is computed. Then the probabilities (ω) of a pixel being in each category (background or object) and the corresponding mean grey level of each category (μ) for every possible threshold value are calculated. Using these values, the inter-class variance is calculated at each possible threshold value using the following formula:

$$\sigma^2(t) = \omega_{bg}(t)\ \omega_{obj}(t)[\mu_{bg}(t)-\mu_{obj}(t)]^2$$

The optimal threshold is then set as the threshold at which $\sigma^2(t)$ is maximized [10].

Next, the minor population of cardiac fibroblasts and other nonmyocytes is removed by filtering out nuclei that do not contain nuclear α-actinin signal (Figure 1-3). Some neonatal cardiac myocytes are binucleated, which causes difficulty for our nucleus-based cell segmentation. Therefore, in the fourth phase adjacent nuclei in the same cell are merged (Figure 1-4). Myocyte nuclei are dilated by a suitable margin width (3 pixels in this case), re-identified, and then eroded by the same margin width to restore their original size.

In the fifth phase (Figure 1-5), cardiac myocyte cell boundaries are segmented and identified. First, cell-background pixel discrimination is conducted based on an Otsu threshold, as previously described. Since many of the cell boundaries are adjacent to one another, additional steps must be taken to segment adjacent myocytes. Then cell-cell boundary segmentation is performed using a previously validated nuclear propagation approach, in which the nucleus of each cell is used as the initial seed and then the algorithm propagates the cell boundary outward [2]. The final boundary between adjacent cells is determined based on an equation using the difference in grey level intensity in the α-actinin channel between adjacent pixels and distance from the nucleus. This segmentation equation approximates the Voronoi segmentation (pixels are assigned to the nearest seed) with an added term accounting for the difference in grey level intensity between adjacent pixels. As a result, the calculated inter-pixel distance is larger in regions of the image with larger gradients in intensity between adjacent pixels. In other words, adjacent pixels with similar intensities are treated as closer together than adjacent pixels with

large differences in grey level intensity. This algorithm is therefore based on the assumption that cell-cell boundaries typically align with large pixel intensity gradients.

In the final phase (Figure 1-6), myocytes touching the edge of the image are removed. This is performed by creating a 1-pixel wide image border and then identifying myocytes that overlap with the border. Nuclei of myocytes touching the image edge are also removed at this stage. Once myocyte segmentation is complete, these objects can be used for subsequent shape and intensity-based measurements.

The automated cell segmentation algorithm was implemented using the open-source MATLAB-based CellProfiler software package [12]. Detailed steps of the CellProfiler pipeline are provided in Table S1. The algorithm and example raw image data from this manuscript are freely available for download at http://bme.virginia.edu/saucerman/.

Texture analysis to automatically quantify sarcomeric organization

We sought to develop a robust quantitative measure of sarcomeric organization using image texture analysis. Several measures of α-actinin image texture available in CellProfiler were tested, including Uniformity, Correlation, Contrast , and Fourier transforms [14]. Haralick [ref 14] is the classic reference for image texture analysis, and these concepts are adapted from his paper. The starting point for texture analysis is the computation of the gray-level co-occurrence matrix from the pixels in a particular segmented cell. The gray-level co-occurrence matrix quantifies the probability that neighboring pixels will have particular gray scale intensity values, and it is computed from the equation:

$$C(i,j) = \sum_{p=1}^{n} \sum_{q=1}^{m} \begin{cases} 1, \text{ if } I(p,q) = i \text{ and } I(p+1,q) = j \\ 0, \text{ otherwise} \end{cases}$$

where *n* and *m* are the dimensions of the image *I*, and *i* and *j* are two particular gray-level values. In MATLAB, the gray-level co-occurrence matrix can be computed using the command "graycomatrix".

Using a normalized version of the gray-level co-occurrence matrix P, there are several ways to quantify various aspects of the image texture. Texture metrics used in this study are Uniformity (closely related to Energy, where Energy = Uniformity$^{0.5}$), Contrast, and Correlation which are given be the following equations from Haralick [14]:

$$Uniformity = \sum_{i,j} P_{ij}^2$$

$$Contrast = \sum_{i,j} P_{ij}(i-j)^2$$

$$Correlation = \sum_{i,j} \frac{(i-\mu_i)(j-\mu_j)P_{ij}}{\sqrt{\sigma_i^2 \sigma_j^2}}$$, where μ is mean, σ is standard deviation

Uniformity quantifies the sum of squares of the co-occurrence matrix, providing an overall measure of the homogeneity of intensities within that segmented cell. Contrast quantifies the degree of difference in pixel intensity from neighboring pixels. Correlation quantifies the linear correlation of a pixel's intensity with its neighbors. As can be seen in Supplementary Figure 4, all three texture measures were correlated with manual scores of sarcomeric organization.

Cluster analysis to identify phenotypic signatures and correlations between phenotypic metrics

Hierarchical cluster analysis was used to identify correlations between hypertrophic agonists and between phenotypic metrics. Briefly, hierarchical clustering aims to identify groups or clusters of related data. A key step in cluster analysis is the selection of an appropriate distance metric to calculate the relatedness of a pair of data. Two distance metrics were used here to compute the distance between two vectors x and y:

$$\text{Euclidean distance: } \sqrt{\sum_i (x_i - y_i)^2}$$

$$\text{Manhattan distance: } \sum_i |x_i - y_i|$$

Cluster analysis was performed using the MATLAB algorithm “clustergram”.