Statistical Analysis Methods

Overview: When the normality of the distribution was rejected, a non-parametric Kruskal-Wallis test was used for comparisons of median values among three or more groups, followed by post-hoc testing using un-paired Mann-Whitney U tests with a Bonferroni-adjusted alpha level. One-way ANOVA with a Tukey's post-hoc multiple comparison test was used to compare more than three means when normality was satisfied. Correlation between variables was determined by using the Spearman's rank correlation test. A mixed effects ANOVA model was used to compare means among repeated, longitudinal measures. A two-sided p-value ≤ 0.05 was considered statistically significant, unless otherwise specified, where a more stringent, or conservative, alpha level than 0.05 was used to define significance due to multiple comparisons. Bar graphs reflect the mean ± standard error of the mean (SEM), while the dot plots include a horizontal line drawn at the median value for each subgroup.

Detailed Methods:

For all of our studies herein, positive and negative controls standardized results on different days for different individuals. A Mann-Whitney U test was used to compare medians between two independent groups, including myocarditis/DCM cases versus controls and for all comparisons involving two levels of a single factor, including Figures 1A, 1C, 2B, 3A-3E, and 8A-8B.

For statistical analysis of the role of gender in Figure 1D and Figures 5A-D, a two-way ANOVA model was fit to investigate male versus female differences for the cases and

controls. A natural log transformation of the response measure was used to satisfy the modeling assumptions of constant variance and normality of the residuals. Noting the limited power to detect a significant interaction, an alpha level of 0.10 was used to identify a significant interaction term. Comparisons were made between males and females after stratifying by case/control status when a significant or marginally-significant interaction was found.

For statistical comparisons in Figure 2A of individuals clinically identified with left or right heart failure (LHF or RHF) or no heart failure (No HF) and Figures 2C-D of individuals with heart failure defined as NYHA Classes I-II cases, NYHA Classes III-IV cases, or normal healthy controls, a Kruskal-Wallis test was used to test for an overall difference in median levels among the independent groups. A significant overall test, at an alpha 0.10 level of significance, was followed by pair-wise testing using a Mann-Whitney test with a Bonferroni-adjusted alpha level of 0.0167.

For studies in Figures 4A-C of the non-recovered subjects, given the small sample size among the cases and the unbalanced design (controls were not measured repeatedly over time), pair-wise comparisons with the control values were made using a Mann-Whitney test and a Bonferroni-adjusted alpha level of 0.0167. No overall test was performed.

For the stimulation of human PBMCs/CD14+monocytes in Figures 6A-D, a mixed effects ANOVA model was fit to account for the correlation among the repeated

measures made on each specimen. The model included a random ID term, as well as a fixed group, treatment and group by treatment interaction. The natural log transformed response measure was analyzed to satisfy modeling assumptions of constant variance and normal residuals. Pairwise testing between cases and controls was performed after stratifying by treatment when a significant interaction was found and multiple comparison adjustment was made using Tukey's method. When considering each treatment separately in Figures 6A-D, we found significant differences after adjustment for multiple comparisons using the Tukey method. In Figure 6A, mean measures were significantly higher for cases compared to controls for TGF-β1 induction by HCM ($p=0.0012$), HCM peptide S2-16 ($p<0.0001$), HCM peptide S2-28 ($p<0.0001$) and peptides S2-16/S2-28 combined ($p<0.0001$). In Figure 6B, after stratifying by treatment and adjusting for multiple comparisons, mean measures were significantly higher for cases compared to controls for IL-6 induction by peptide S2-28 ($p=0.017$) and peptides S216/S2-28 ($p=0.0010$). After stratifying by treatment and with adjustment for multiple comparisons in Figure 6C, mean measures were significantly higher for cases compared to controls for IL-23 induction by peptide S2-16 ($p=0.026$), peptide S2-28 ($p=0.0005$) and S2-16/S2-28 combined ($p=0.0032$). In Figure 6D, mean measures were significantly higher for cases compared to controls for IL-1β induction by HCM peptide S2-16 ($p=0.0003$), HCM peptide S2-28 ($p<0.0001$) and peptides S2-16/S2-28 combined ($p=0.0004$).

For the TLR 2 blocking experiment in Figures 7A-C, a mixed effects ANOVA model was fit to account for the correlation among the repeated measures made on each specimen

(n=3). The model included a random sample term as well as a fixed treatment term. The natural log transformed response measure was analyzed to satisfy modeling assumptions of constant variance and normal residuals. Pairwise testing between pairs of treatments was performed using Tukey's method to adjust for multiple comparisons.

Detailed Statistical Analysis for Figures 7A-C: Figure 7A: Based on alpha levels that were adjusted for multiple comparisons using Tukey's method, all pairwise comparisons with α-TLR2 only, α-TLR2+S-216, and α-TLR2+S2-28 are statistically significant ($p<0.006$ for each) as are the comparisons between +S2-16 versus LPS or media ($p<0.02$ for each) and the comparison between +S2-28 and LPS ($p=0.026$). No other comparisons were statistically significant. Figure 7B: Based on alpha levels that were adjusted for multiple comparisons using Tukey's method, all pairwise comparisons with α-TLR2 only, α-TLR2+S2-16, and α-TLR2+S2-28 are statistically significant ($p<0.001$ for each), with the only exceptions being comparisons among these three treatments ($p>0.9$ for each) and comparisons between these three treatments and Media ($p>0.1$ for each). In addition, all other pairwise comparisons with Media were significant ($p<0.0001$ for each) as were comparisons between +S2-16 and LPS ($p=0.0016$). No other comparisons were statistically significant. Figure 7C: Based on alpha levels that were adjusted for multiple comparisons using Tukey's method, all pairwise comparisons with Isotype antibody/ α-TLR2 only, Isotype antibody/ α-TLR2 +S2-16, and Isotype antibody/α-TLR2+S2-28 are statistically significant ($p<0.02$ for each), with the only exceptions being comparisons among these three treatments ($p>0.2$ for each) and comparisons among +S2-16, +S2-28, Isotype antibody/ α-TLR2 + S2-16, and α-TLR2 +

S2-28 (p>0.5 for each). In addition, comparisons between +S2-16 or +S2-28 and α-TLR2 only, α-TLR2+S2-16, and α-TLR2 + S2-28 are statistically significant (p<0.001 for each). Finally, LPS and Media differ from +S2-16 and +S2-28 (p<0.001 for each). No other comparisons were statistically significant. In addition, mean IL-6 and TGF□□ measures in Figure 7 significantly differed from controls and were significantly reduced among anti-TLR2 treatment groups (p<0.0001, ANOVA overall test). A mixed effects ANOVA model was fit to account for the correlation among the repeated measures made on each specimen (n=3). Pairwise testing between pairs of treatments was performed using Tukey's method to adjust for multiple comparisons.

Spearman's rank correlation coefficient was used to summarize the strength of the linear association between continuous measures in Figure 9.

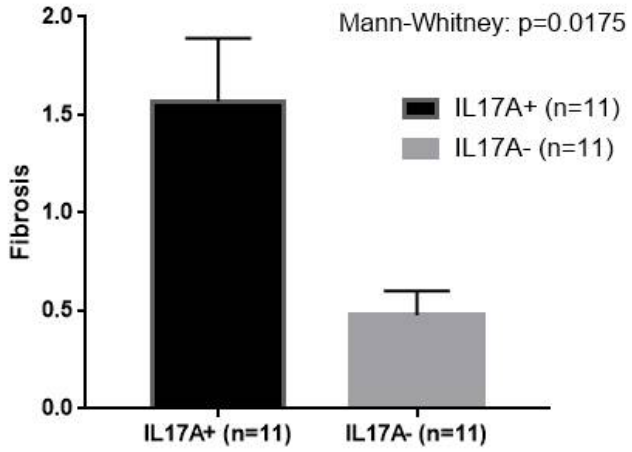
Trichrome staining of heart biopsies

Tissue sections were stained using Masson's Trichrome Method (90) (AFIP modification). Briefly, tissues were de-paraffinized and hydrated in distilled water, then re-fixed in Bouin's solution [saturated picric acid, formaldehyde (37-40%), glacial acetic acid] for 1 hour at 56° C, or overnight at room temperature when formalin fixed. Tissues were cooled and washed in running water to remove yellow color, then rinsed in distilled water. Sections were then stained in Wiegert's iron hematoxylin solution [stock solution A: hematoxylin and 95% alcohol; stock solution B: 29% ferric chloride in water, distilled water, concentrated hydrochloric acid; working solution - equal parts of stock solutions A and B] for 10 minutes, washed in running water for 10 minutes, and rinsed in distilled

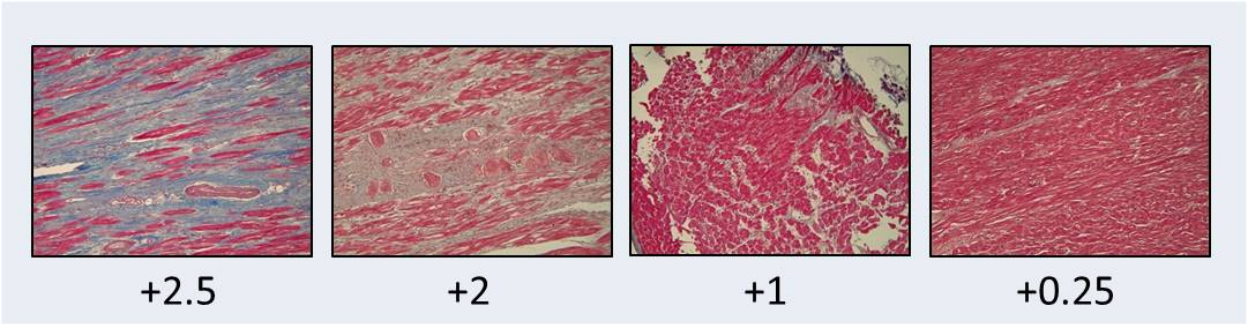
water. Tissues were then stained in Biebrich scarlet-acid fuchsin solution (Biebrich scarlet, 1% aqueous, acid fuchsin, 1% aqueous, and glacial acetic acid) for 3 minutes, then rinsed in distilled water. Tissues were differentiated in phosphomolybdic-phosphotungstic acid solution (5% phosphomolybdic acid, 5% phosphotungstic acid) for 10 minutes. Solution was discarded, and tissue sections were stained with aniline blue solution (aniline blue, glacial acetic acid, distilled water) for 5 minutes. Sections were rinsed in distilled water, placed in 1% glacial acetic acid solution for 3 minutes, then dehydrated in 95% ethyl alcohol, absolute ethyl alcohol, and cleared in xylene, two changes each. Sections were mounted with Permount or Histoclad. In trichrome-stained tissue sections, nuclei stain black and collagen stains blue. Cytoplasm, keratin, muscle fibers and intercellular fibers stain red.

Supplementary Figures

A



B



Supplementary Figure 1: Fibrosis in IL17A+ vs IL17A- Myocardial Biopsies. (A) Fibrosis scores were determined using trichrome staining (graded +0.25 to +3) and were compared between IL17A+ (N=11) and IL17- (N=11) biopsies. Trichrome staining revealed that IL-17A+ biopsies had a trend toward heavy fibrosis while those which were IL-17A- were less fibrotic. Mann-Whitney: p= 0.0175. **(B)** Representative trichrome grading as observed at 100X (10X objective X 10X eyepiece). +2.5 trichrome score shows moderate to marked interstitial fibrosis, whereas, +0.25 trichrome score shows very mild interstitial fibrosis.