

## Supplementary Information

for

### **Patients with ACPA-positive and ACPA-negative Rheumatoid Arthritis Show Different Serological Autoantibody Repertoires and Autoantibody Associations with Disease Activity**

Kevin Y. Cunningham<sup>1</sup>, Benjamin Hur<sup>2,3</sup>, Vinod K. Gupta<sup>2,3</sup>, Courtney A. Arment<sup>4</sup>, Kerry A. Wright<sup>4</sup>, Thomas G. Mason<sup>4</sup>, Lynne S. Peterson<sup>4</sup>, Delamo I. Bekele<sup>4</sup>, Daniel E. Schaffer<sup>4</sup>, Marissa L. Bailey<sup>4</sup>, Kara E. Delger<sup>4</sup>, Cynthia S. Crowson<sup>4,5</sup>, Elena Myasoedova<sup>4,5</sup>, Hu Zeng<sup>4,6</sup>, Moses Rodriguez<sup>7</sup>, Cornelia M. Weyand<sup>4,6</sup>, John M. Davis III<sup>4,†</sup>, and Jaeyun Sung<sup>1,2,3,4,†,\*</sup>

<sup>1</sup>Bioinformatics and Computational Biology Program, University of Minnesota Twin Cities, Minneapolis, MN 55455, USA

<sup>2</sup>Microbiome Program, Center for Individualized Medicine, Mayo Clinic, Rochester, MN 55905, USA

<sup>3</sup>Division of Surgery Research, Department of Surgery, Mayo Clinic, Rochester, MN 55905, USA

<sup>4</sup>Division of Rheumatology, Department of Medicine, Mayo Clinic, Rochester, MN 55905, USA

<sup>5</sup>Department of Quantitative Health Sciences, Mayo Clinic, Rochester, MN 55905, USA

<sup>6</sup>Department of Immunology, Mayo Clinic, Rochester, MN 55905, USA

<sup>7</sup>Department of Neurology, Mayo Clinic, Rochester, MN 55905, USA

<sup>†</sup>John M. Davis III and Jaeyun Sung contributed equally as co-senior authors

\*Correspondence: [Sung.Jaeyun@mayo.edu](mailto:Sung.Jaeyun@mayo.edu)

## Supplementary Files

**Supplementary Figure 1.** Autoantibodies targeting CISH proteins are positively correlated with Clinical Disease Activity Index (CDAI).

**Supplementary Table 1.** Clinical and demographic data for patients with ACPA+ RA (n = 32) and ACPA– RA (n = 30).

**Supplementary Table 2.** Demographic data for healthy controls.

**Supplementary Table 3.** Means of the quantile normalized autoantibody abundances in all three study groups.

**Supplementary Table 4.** Means of quantile normalized autoantibody abundances with corresponding fold changes.

**Supplementary Table 5.** Differentially abundant autoantibodies higher in RA subgroups compared with controls.

**Supplementary Table 6.** Differentially abundant autoantibodies lower in RA subgroups compared with controls.

**Supplementary Table 7.** Differentially abundant autoantibodies between ACPA-positive RA and ACPA-negative RA.

**Supplementary Table 8.** Enriched GO Biological Processes for autoantigens whose autoantibodies are higher in ACPA+ RA or ACPA– RA (compared with healthy controls).

**Supplementary Table 9.** Enriched GO Biological Processes for autoantigens whose autoantibodies are lower in ACPA+ RA or ACPA– RA (compared with healthy controls).

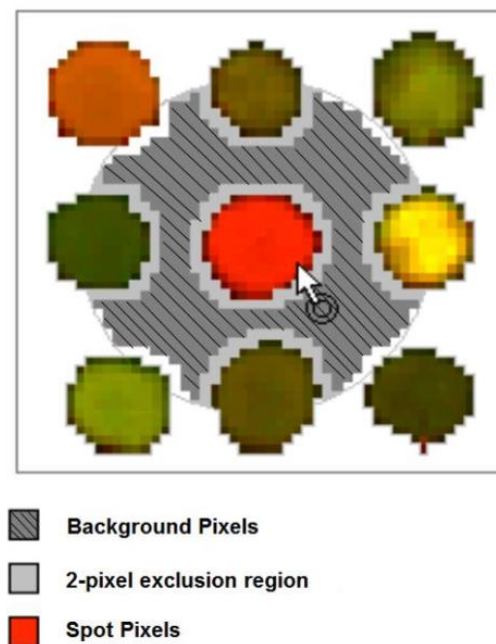
**Supplementary Table 10.** Spearman correlation coefficients between autoantibody abundances and Clinical Disease Activity Index.

**Supplementary Table 11.** Raw abundance values of autoantibodies (RFUs) across all serum samples.

## Microarray data handling and pre-processing

For each slide, proteins and control probes are in quadruplicates, with 4 arrays on each slide.

**Step 1:** Calculate the net intensities for each spot by subtracting the background signal intensities from the foreground signal intensities of each spot. For each spot, the background signal intensity is calculated using a circular region that is centered on the spot. The figure below is a visual representation of how this calculation was carried out (*Molecular Devices*). The red spot in the middle is the spot of interest.



**Step 2:** Remove replica spots with an RFU  $\leq 0$ .

**Step 3:** Zero net intensities if only 1 replica spot remaining.

**Step 4:** Calculate the percentage of coefficient of variant (CV%) to determine the variations between the replica spots on each slide.

$$CV\% = \frac{S.D.}{Mean} \times 100\%$$

Flag a set of replica spots with only 2 or less replica/s remaining and  $CV\% > 20\%$  as “High CV”. The mean RFU of these replica spots (proteins) will be excluded from the downstream analysis.

For the proteins and controls with a  $CV\% > 20\%$  and with 3 or more replica spots remaining, the replica spots that resulted in this high CV% value were filtered out. This was done by first calculating the standard deviation between the median value of the net intensities and the individual net intensities for each set of the replica spots. The spot with the largest standard deviation was removed. CV% values were re-calculated and the process repeated.

**Step 5:** Calculate the mean of the net intensities for the remaining replica spots.

**Step 6:** Inspect the signal intensities of the two positive controls: IgG and Cy3-BSA.

**Step 7:** Composite normalization of data using both quantile-based and total intensity-based modules (Duarte, J. *et al.*, 2013). This method assumes that the different samples share a common distribution of their control probes while taking into account the potential existence of flagged spots within them. The Immunome array uses Cy3-labeled biotinylated BSA (Cy3-BSA) replicates as the positive control spots across slides. It is considered a housekeeping probe for normalization of signal intensities for any given study.

The quantile module adopts the algorithm in Bolstad *et al.*, 2003. This enables the detection and handling of outliers or flagged spots in any of the Cy3-BSA control probes. A total intensity-based module was then implemented to obtain a scaling factor for each sample. This assumes that post-normalization, the positive controls should have a common total intensity value across all samples (Causton H.C. *et al.*, 2004). This method aims to normalize the protein array data from variations in their measurements while preserving the targeted biological activity across samples.

#### **Quantile-based normalization of all Cy3-BSA across all samples**

(i = spot number and j = sample number)

1. Load all Cy3-BSA across all samples, j, into an  $i \times j$  matrix X
2. Sort spot intensities in each column j of X to get  $X_{\text{sort}}$
3. Take the mean across each row i of  $X_{\text{sort}}$  to get  $\langle X_i \rangle$

#### **Intensity-based normalization**

1. Calculate the sum of the mean across each row i,  $\sum \langle X_i \rangle$
2. For each sample, k, calculate the sum of all Cy3-BSA controls,  $\sum X_k$
3. For each sample, k,

$$\text{Scaling factor}(k) = \frac{\sum \langle X_i \rangle}{\sum X_k}$$

## **References**

1. Bolstad, B. M., Irizarry, R. A., Astrand, M. & Speed, T. P. (2003). A Comparison of Normalization Methods for High Density Oligonucleotide Array Data Based on Bias and Variance. *Bioinformatics*. 19(2):185-193.
2. Causton, H.C., Quackenbush, J. & Brazma, A. (2004). *Microarray gene expression data analysis: a beginners' guide*. 1st ed. Malden: Blackwell Publishing.
3. Duarte, J. et al. (2013). Protein function microarrays: design, use, and bioinformatic analysis in cancer biomarker discovery and quantitation. *Bioinformatics of Human Proteomics*, Springer Netherlands. 39-74.
4. Molecular Devices, GenePix Pro 7 Software User Guide, 0112-0178 Rev B (June 2010) Figure 1-1 p25.