Biostatistics (2017), **0**, 0, *pp.* 1–13
doi:10.1093/biostatistics/gel-supplementary-materials

Supplementary Materials to "Estimating AutoAntibody Signatures to Detect Autoimmune Disease Patient Subsets"

ZHENKE WU*[,1], LIVIA CASCIOLA-ROSEN[2], AMI A. SHAH[2],

ANTONY ROSEN[2], SCOTT L. ZEGER[3]

[1] *Department of Biostatistics and Michigan Institute of Data Science, University of Michigan, Ann Arbor, Michigan 48109*

[2] *Division of Rheumatology, Department of Medicine, Johns Hopkins University School of Medicine, Baltimore, Maryland, 21224*

[3] *Department of Biostatistics, Johns Hopkins University, Baltimore, MD 21205*

*zhenkewu@umich.edu

Appendix S1. Raw Data and Standardization

Let $(\mathbf{t}^{\mathsf{raw}}, \mathbf{M}^{\mathsf{raw}}) = \left\{(t_{gs}^{\mathsf{raw}}, M_{gis}^{\mathsf{raw}})\right\}$ represent the raw, high-frequency GEA data, for pixel $s = 1, \ldots, S_g$ on lane $i = 1, \ldots, N_g$ from gel $g = 1, \ldots, G$. Here $\mathbf{t}^{\mathsf{raw}}$ is a grid that evenly splits the unit interval $[0, 1]$ with $t_{gs}^{\mathsf{raw}} = s/S_g \in [0, 1]$, where a large S_g represents a high imaging resolution. Note that in the raw data, S_g varies by gel from $1,437$-$1,522$ in our application. M_{gis}^{raw} is the radioactive intensity scanned at t_{gs}^{raw} for lane $i = 1, \ldots, N_g$ on Gel $g = 1, \ldots, G$. Let $N = \sum_g N_g$ be the total number of samples tested.

For the rest of this section, we process the high-frequency data $(\mathbf{t}^{\mathsf{raw}}, \mathbf{M}^{\mathsf{raw}})$ into high-frequency data $(\mathbf{t}^0, \mathbf{M}^0)$ that have been standardized across multiple gels. The latter will be used as input for peak detection (Section 2.2) and batch effects correction (Section 2.3).

*To whom correspondence should be addressed.

i. **Smoothing**. For each sample lane, smooth the raw intensity data by LOESS, with a span $h = 0.022$. Let $\widetilde{\mathbf{M}} = \{\widetilde{M}_{gis}\}$ denote the smoothed mean function evaluated at raw imaging location t^{raw}_{gs}.

ii. **Standardization Across Gels**. Imaging resolution may vary by gel, we hence standardize the smoothed intensity values $\widetilde{\mathbf{M}}$ into $B^0 = 700$ bins using a set of evenly-spaced break points $\{0 = \kappa_0 < \kappa_1 < \ldots < \kappa_{B^0} = 1\}$ shared by all gels.

We clip the images at the right end $\{b : t_b > 0.956\}$ because they represent small molecular weight molecules migrating at the dye front (that is, not separable by gel type used). Their exclusion from autoantibody comparisons is standard practice. We denote the standardized data by $(\mathbf{t}^0, \mathbf{M}^0) = \left\{\left(t^0_b, M^0_{gib}\right)\right\}$.

Appendix S2. Peak Calling Algorithm

Given a half-width h, collect the *peak-candidate bins* defined by $\mathcal{B}^0_{gi}(h) = \{b \mid \mathsf{score}_{gi}(b) = \nu\}$. Because h controls the locality of a peak, we perform peak-candidacy search for a few values of h. The union of the identified peak-candidate bins under various h, $\mathcal{B}^0_{gi} = \cup_h \mathcal{B}^0_{gi}(h)$, generally are comprised of multiple blocks, one per set of contiguous peak-candidate bins. Among the blocks, we merge two nearby ones, for example, if $\mathcal{B}^0_{gij}$ and $\mathcal{B}^0_{gi,j+1}$ satisfy $\min \mathcal{B}^0_{gi,j+1} - \max \mathcal{B}^0_{gij} \leqslant \gamma(= 5)$. We also remove short blocks of length less than three to obtain the final peak-candidate bins $\{\mathcal{B}_{gij}\}_{j=1}^{J_{gi}}$. Finally, we pick the bin b_{gij} that maximizes its within-block intensities and denote them *peak* $j = 1, \ldots, J_{gi}$ for lane $i = 1, \ldots, N_g$ on gel $g = 1 \ldots, G$.

The true and false peak detection rates are determined by several factors including the half-peak width h, the minimum intensity elevation C_0, the true intensities at each bin and the measurement errors inherent in autoradiography. We calibrate the first two parameters so that 1) the reference lanes have exactly 7 detected peaks (perfect observed sensitivity and specificity), and 2) the peaks for actin stand out clearly. For example, middle panel of Figure 1(b) shows the

result of peak detection by blue asterisks for one set of the gels, where the peaks rising slightly above the background are effectively captured, most notably for lanes 5,10,15 where the small actin peaks are identified. Note that, we have reduced the impact of measurement noise on peak detection by computing the local difference scores from the smoothed data rather than the raw data. In our analyses, we have chosen the minimum peak amplitude parameter $C_0 = 0.01$ of higher order than the noise level obtained from LOESS smoothing.

Alternative approaches to peak detection include random process modeling (e.g., Carlson *and others*, 2015), multiplicity adjustment after local maxima hunting (e.g., Schwartzman *and others*, 2011) and filtering methods (e.g., Du *and others*, 2006). From our experience, they are designed and hence more suitable for data with appreciably higher noise levels; our data show much lower noise level in the measured autoradiographic intensities. For example, in random process models that are motivated by the analysis of pulsatile, or episodic time series data, the unknown locations of peaks and the observed intensity values are modeled by double stochastic processes, such as Cox processes, to fit the continuous intensity data for each gel and sample (e.g., Carlson *and others*, 2015). However, because a tiny peak has a narrow span, its associated few observed data is not as informative about the peak location compared to that for a wider peak, hence tends to identify small peaks with larger uncertainties in peak presence/absence and, if present, its location. In addition, fitting the random process models for peak detection for hundreds of subjects and hundreds of dimensions per person involves iterative MCMC sampling and could be computationally expensive.

Appendix S3. Reference Alignment via Piecewise Linear Dewarping

We align all the gels towards an arbitrarily chosen template gel g_0 using piecewise linear dewarping (Uchida and Sakoe, 2001) whose knots are anchored at gel-specific marker peaks $\left\{\mathcal{P}^0_{g1}\right\}_1^G$. We first match the marker peaks $\mathcal{P}^0_{g1}$ observed on a query gel g to the reference peaks $\mathcal{P}^0_{g_0 1}$ on the

template gel g_0, and then linearly stretch or compress gel g between the reference peaks.

Let the function $\mathcal{W}_g(\cdot; g_0) : b \mapsto b'$ denote the matching of the b-th bin of the template gel g_0 to the b'-th bin of gel g to be dewarped. Suppose in the template gel g_0 the b-th bin falls within two neighboring reference peaks $T_{g_0 1j} \leqslant t_b < T_{g_0 1,j+1}$, where $j = j(b; g_0)$ is determined by the bin number b and the gel g_0 used as reference. Let

$$\mathcal{W}_g(b; g_0) = \lfloor w \cdot b_{g1,j+1} + (1-w) b_{g1j} \rfloor, \tag{A1}$$

where $w = w(b; g_0) = (b - b_{g_0 1j}) \,/\, (b_{g_0 1,j+1} - b_{g_0 1,j})$, and $\lfloor a \rfloor$ is the largest integer smaller than or equal to a. The piecewise linear dewarping function for gel $g \neq g_0$, defined as

$$\sum_j \mathcal{W}_g(b; g_0) \mathbf{1}_{\{T_{g_0 1j} \leqslant t_b < T_{g_0 1,j+1}\}},$$

corrects the batch effects for all the lanes by anchoring gel stretching or compression at the locations representing the marker molecules $\{\mathcal{P}^0_{g1}\}_1^G$.

Figure Figure S1 illustrates the results before and after batch-effect correction. The piecewise linear dewarping automatically matches the marker peaks from multiple gels to facilitate cross-gel band comparisons.

Appendix S4. Details on Posterior Sampling Algorithm

We sample from the joint posterior by the following algorithm:

1. Update peak-to-landmark indicators Z_{gij} for peak $j = 1, \ldots, J_{gi}$, lane $i = 1, \ldots, N_g$ and gel $g = 1, \ldots, G$ by categorical distribution

$$\mathbb{P}(Z_{gij} = \ell \mid others) \propto N\left(T_{gij}; \mathcal{S}_g(u_i, \nu_\ell; \boldsymbol{\beta}_g), \sigma^2_\epsilon\right) \{1 - \exp(-\lambda^*_\ell)\}, \tag{A2}$$

   for $\ell = 1, \ldots, L$ that satisfy the support constraint $|\nu_\ell - T_{gij}| < A_0$.

2. Update the B-spline basis coefficients $\boldsymbol{\beta_g} = [\beta_{gst}]$ for gel $g = 1, \ldots, G$. Let Δ_1 be the first order difference operator of dimension $(T_\nu - 1) \times T_\nu$ with entries $\Delta_{1ij} = \delta(i+1, j) - \delta(i, j)$

and $\delta(i,j)$ is the (i,j)th entry in identity matrix $\mathbf{I}_{T_\nu}$; Similarly let Δ_2 with $\Delta_{2ij} = \delta(i+1,j) - \delta(i,j)$ with δ_{ij} from $\mathbf{I}_{T_u}$. The random walk priors (2.6) and (2.7) can be written as $\boldsymbol{\beta}_{gs\cdot} \overset{d}{\sim} N_{T_\nu - 1}(\cdot; \boldsymbol{\beta}^{id}_{[-T_\nu]}, \sigma_{g1}^{-2}\Delta_1'\Delta_1)\mathbf{1}\{\nu_0 = \beta_{g11} <, \ldots, \beta_{g1,T_\nu-1} < \nu_{L+1}\}$ and $\boldsymbol{\beta}_{g\cdot t} \overset{d}{\sim} N_{T_u}(\cdot; \mathbf{0}, \sigma_{gt}^{-2}\Delta_2'\Delta_2)$. Although both Δ_1 and Δ_2 are rank deficient, we show below that the conditional posterior for $\boldsymbol{\beta}_g$ is proper under *scattering condition*.

Update the B-spline basis coefficients $\boldsymbol{\beta_g} = [\beta_{gst}]$ for gel $g = 1, \ldots, G$ by multivariate normal distribution

$$\begin{aligned}[\mathsf{vec}\{\boldsymbol{\beta}_g'\} \mid others] \propto & \exp(-\frac{1}{\sigma^2}\|\boldsymbol{T}_g - \boldsymbol{B}_g\mathsf{vec}[\boldsymbol{\beta}_g']\|_2^2) \\ & \times \exp(-\frac{1}{\sigma_{g1}^2}\|\Delta_1^{\mathsf{aug}}\boldsymbol{\beta}^{\mathsf{id,aug}} - \Delta_1^{\mathsf{aug}}\mathsf{vec}[\boldsymbol{\beta}_g']\|_2^2) \\ & \times \exp(-\sum_{t=2}^{T_\nu}\frac{1}{\sigma_{gt}^2}\|\Delta_{2,t}^{aug}\mathsf{vec}[\boldsymbol{\beta}_g']\|_2^2),\end{aligned}$$

where $\boldsymbol{\beta}^{\mathsf{id,aug}}$ is a column vector formed by stacking $\boldsymbol{\beta}^{\mathsf{id}}$ T_u times hence of length T_uT_ν; Δ_1^{aug} is a matrix augmenting Δ_1 to $[\Delta_1 \mid \mathbf{0}_{(T_\nu-1)\times(T_uT_\nu - T_\nu)}]$; $\Delta_{2,t}^{\mathsf{aug}}$ augments Δ_2 to a $(T_u - 1) \times T_uT_\nu$ matrix whose $(t, t+T_\nu, \ldots, t+(T_u-1)T_\nu))$-th columns correspond to $\beta_{g1t}, \beta_{g2t}, \ldots, \beta_{g,T_u,t}$ and form a submatrix identical to Δ_2.

The conditional distribution above simplifies to a multivariate normal distribution with mean vector

$$\boldsymbol{\Lambda}_g^{-1}\left\{\sigma_\epsilon^{-2}\boldsymbol{B}_g'\boldsymbol{T}_g + \sigma_{g1}^{-2}\Delta_1^{\mathsf{aug}'}\Delta_1^{\mathsf{aug}}\boldsymbol{\beta}^{\mathsf{id,aug}}\right\}$$

where precision matrix $\boldsymbol{\Lambda}_g = \sigma_\epsilon^{-2}\boldsymbol{B}_g'\boldsymbol{B}_g + \sigma_{g1}^{-2}\Delta_1^{\mathsf{aug}'}\Delta_1^{\mathsf{aug}} + \sum_{t=2}^{T_\nu}\sigma_{gt}^{-2}\Delta_{2,t}^{\mathsf{aug}'}\Delta_{2,t}^{\mathsf{aug}}$. Given large σ_{gt}^2, the mean vector will be close to $\boldsymbol{\beta}^{\mathsf{id,aug}}$, i.e. no warping, if the smoothness parameter σ_{g1}^{-2} is large and otherwise close to the flexible fitted surface by the observed peaks. Turning to the smoothing parameter in the u-direction, σ_{gt}^{-2}, the Gamma-InversePareto mixture (Appendix S5) encourages the sampling chains to jump between no versus flexible warping in the u-direction.

The matrix $\boldsymbol{B}_g'\boldsymbol{B}_g$ is full rank when the observed peaks are well scattered across lanes and along the gels. Let (c_{t_0}, c_{t_0+1}) be the support of the t_0-th B-spline basis $B_{2t_0}(\cdot)$ in the ν-direction. Suppose, for example, no peaks appear in (c_{t_0}, c_{t_0+1}). $\boldsymbol{B}_g$ will be constant zeros for columns t_0, $t_0 + T_\nu$,..., and $t_0 + T_\nu(T_u - 1)$ thus rank deficient. As a result, the posterior of $\{\beta_{gst_0}\}_{s=1}^{T_u}$ will not converge to a point mass given infinite samples and can only be learned through its neighboring coefficients via random walk priors (2.6) and (2.7). Rank deficiency also occurs if multiple neighboring lanes have no observed peaks. Given sparse peaks, though $\boldsymbol{B}_g$ can be made full rank by reducing the number of basis functions, judicious trade-off between flexibility of $\mathcal{S}_g$ and parameter identifiability is necessary for specific applications. We refer to the condition that $\boldsymbol{B}_g$ is full rank as *scattering condition*.

In our applications, failure of the scattering condition is rare. $\boldsymbol{\Lambda}_g$ is then a sum of positive definite matrix and semi-definite matrices and hence is invertable. Also recall that $\boldsymbol{B}_g'\boldsymbol{B}_g$ is sparse as constructed from sparse B-spline bases. Because $\Delta_1^{\mathsf{aug}'}\Delta_1^{\mathsf{aug}}$ and $\sum_{t=1}^{T_\nu}\sigma_{gt}^{-2}\Delta_2^{\mathsf{aug}'}\Delta_2^{\mathsf{aug}}$ are both sparse square matrix with at most $\mathcal{O}(T_uT_\nu)$ nonzeros, $\boldsymbol{\Lambda}_g$ preserves the sparsity of $\boldsymbol{B}_g'\boldsymbol{B}_g$. So we use sparse Cholesky factorization of $\boldsymbol{\Lambda}_g$ to produce its Cholesky factors.

We first block update $\{\beta_{gst}\}_{s=1}^{T_u}$ for $t = 2$ from $[\beta_{gst} \mid \beta_{gj_1j_2}, j_2 \neq t, others]$ with constraint $\beta_{gst} > \beta_{gs1} = \nu_0$, $s = 1,\ldots,T_u$ and continue for $t = 3,\ldots,T_\nu - 1$. This step requires calculation of inverse of submatrices of $\boldsymbol{\Lambda}_g^{-1}$ for T_ν times. In our application, computing one such inverse when $T_u = 6$ and $T_\nu = 10$ requires < 0.001 seconds.

3. Update the smoothing parameters $\tau_{gt}^2 = \sigma_{gt}^{-2}$ and smoothness selectio indicator ξ_{gt} (Appendix S5). First randomly switch ξ_{gt} to ξ_{gt}^* either from 0 to 1 or 1 to 0 for $t = 1,\ldots,T_\nu$, $g = 1,\ldots,G$. For the parameter τ_{gt}^2, we propose its candidate τ_{gt}^{*2} from the log-normal distribution with log-mean parameter τ_{gt}^2. We accept $(\tau_{gt}^{*2}, \xi_{gt}^*)$ with probability

$$\alpha_g^{(1)} = \min\left\{1, \frac{p(\boldsymbol{T}_g;\boldsymbol{\beta}_g,\tau_{gt}^{*2})\pi(\tau_{gt}^{*2} \mid \xi_{gt}^*)q(\tau_{gt}^2 \mid \tau_{gt}^{*2})q(\xi_{gt} \mid \xi_{gt}^*)}{p(\boldsymbol{T}_g;\boldsymbol{\beta}_g,\tau_{gt}^2)\pi(\tau_{gt}^2 \mid \xi_{gt})q(\tau_{gt}^2 \mid \tau_{gt}^2)q(\xi_{gt}^* \mid \xi_{gt})}\right\}. \tag{A3}$$

We update τ_{gt}^2 again because it is continuous and therefore has a much bigger parameters space than that of discrete parameter. Using random walk Metroplis-within-Gibbs algorithm, we propose τ_{gt}^{*2} from the log-normal distribution with log-mean parameter τ_{gt}^2 and accept with probability

$$\alpha_g^{(2)} = \min\left\{1, \frac{p(\boldsymbol{T}_g; \boldsymbol{\beta}_g, \tau_{gt}^{*2})\pi(\tau_{gt}^{*2} \mid \xi_{gt}^*)q(\tau_{gt}^2 \mid \tau_{gt}^{*2})}{p(\boldsymbol{T}_g; \boldsymbol{\beta}_g, \tau_{gt}^2)\pi(\tau_{gt}^2 \mid \xi_{gt})q(\tau_{gt}^{*2} \mid \tau_{gt}^2)}\right\}. \tag{A4}$$

4. Update the smoothing parameter in the ν-direction $\tau_{g1}^2 = \sigma_{g1}^{-2}$, $g = 1, \ldots, G$ by proposing its candidate τ_{g1}^2 from the log-normal distribution with log-mean parameter τ_{g1}^2. We accept the proposal with probability $\min\left\{1, \frac{N_{T_\nu-1}\left(\{\beta_{g1t}\}_{t=1}^{T_\nu-1}; \mathbf{0}, \tau_{g1}^{*-2}\mathbf{I}\right)q(\tau_{g1}^2|\tau_{g1}^{*2})}{N_{T_\nu-1}\left(\{\beta_{g1t}\}_{t=1}^{T_\nu-1}; \mathbf{0}, \tau_{g1}^{-2}\mathbf{I}\right)q(\tau_{g1}^{*2}|\tau_{g1}^2)}\right\}$.

5. Update smoothness selection hyperparameter ρ_g (Appendix S5), for $g = 1, \ldots, G$ from

$$[\rho_g \mid others] \sim \mathsf{Beta}(a_\rho + \sum_t \mathbf{1}\{\xi_{gt} = 1\}, b_\rho + \sum_t \mathbf{1}\{\xi_{gt} = 0\}). \tag{A5}$$

We calibrate the scale of the proposals in Step 3-4 at the burn-in period of the MCMC to achieve an acceptance rate between 30% and 70%.

Appendix S5. Details on Shrinkage Prior for Hyperparameters $\{\sigma_{gt}^{-2}\}$

For the hyperpriors on the smoothing parameter for the warping function $\mathcal{S}_g$ in the u-direction, $\{\tau_{gt}^2 = \sigma_{gt}^{-2}\}$, we specify a mixture with two well-separated component distributions with one favoring small and the other large values (Morrissey *and others*, 2011):

$$\tau_{gt}^2 \sim \xi_{gt}\mathsf{Gamma}(\cdot \mid a_\tau, b_\tau) + (1 - \xi_{gt})\mathsf{InvPareto}(\cdot \mid a'_\tau, b'_\tau), \tag{A6}$$

$$\mathsf{InvPareto}(\tau; a, b) = \frac{a}{b}\left(\frac{\tau}{b}\right)^{a-1}, a > 0, 0 < \tau < b, \tag{A7}$$

where the Gamma-distributed component ($a_\tau = 3$, $b_\tau = 2$) concentrates near smaller values while the inverse-Pareto component prefers larger values ($a'_\tau = 1.5$, $b'_\tau = 400$).

We designed the mixture priors for τ_{gt}^2 so that given t, β_{gst}, $s = 1, \ldots, T_u$ are similar or discrepant according as τ_{gt}^2 being large or small. This bimodal mixture distribution creates a sharp

separation between flexible and smooth warping functions across lanes on a gel (u-direction). The random smoothness indicator ξ_t represents a flexible (1) or constant (0) relationship of warping across lanes. We let $\xi_{gt} \sim \mathsf{Bernoulli}(\rho_g)$ with success probability ρ_g and then put a hyperprior $\rho_g \sim \mathsf{Beta}(a_\rho, b_\rho)$ to let data inform the degree of smoothness. In this paper's application, we use $a_\rho = b_\rho = 1$ so that the prior has a mean of $1/2$ that assigns equal prior probabilities to all submodels with flexible or constant warping functions across lanes; in the case of high-dimensional basis function (large T_ν), the Beta prior with other hyperparameters can also allow a prior that lets the fraction of constant warpings $\rho_g = \rho_{gp}$ to approach 1 as $p \to \infty$.

We specify the hyperprior for the smoothing parameter $\tau_{g1}^2 = \sigma_{g1}^{-2} \overset{d}{\sim} \mathsf{InvPareto}(\cdot \mid a'_\tau, b'_\tau)$. We also fix the measurement error variance $\sigma_\epsilon = \Delta/3$ where Δ is the minimum distance among grid points $\{\nu_\ell\}$ in the standardized scale. These parameters are chosen to constrain the shape of $\mathcal{S}_g$ and are shown to have good dewarping performances, e.g., aligning all actin peaks towards a single landmark.

Appendix Figures

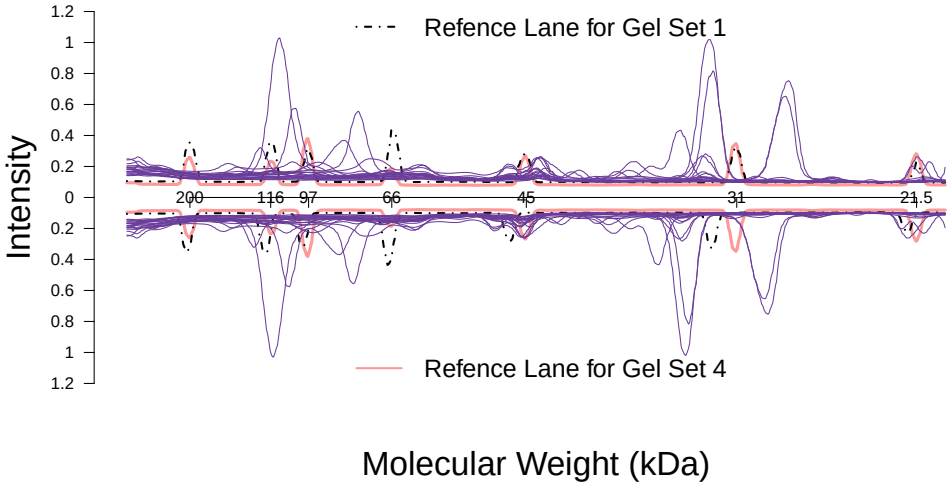

Figure S1. Before (bottom) and after (top) piecewise linear dewarping towards Gel 4. *Top*: 21 intensity curves, 20 solid curves from one GEA experiement ($g = 1$) after reference alignment; one dashed curve for the reference lane in gel $g_0 = 4$. The two curves *not* in purple denote the Lane 1 intensities from the two gels and are aligned. *Bottom*: The same 21 intensity curves without reference alignment. The reference lanes (non-purple ones) are mismatched.

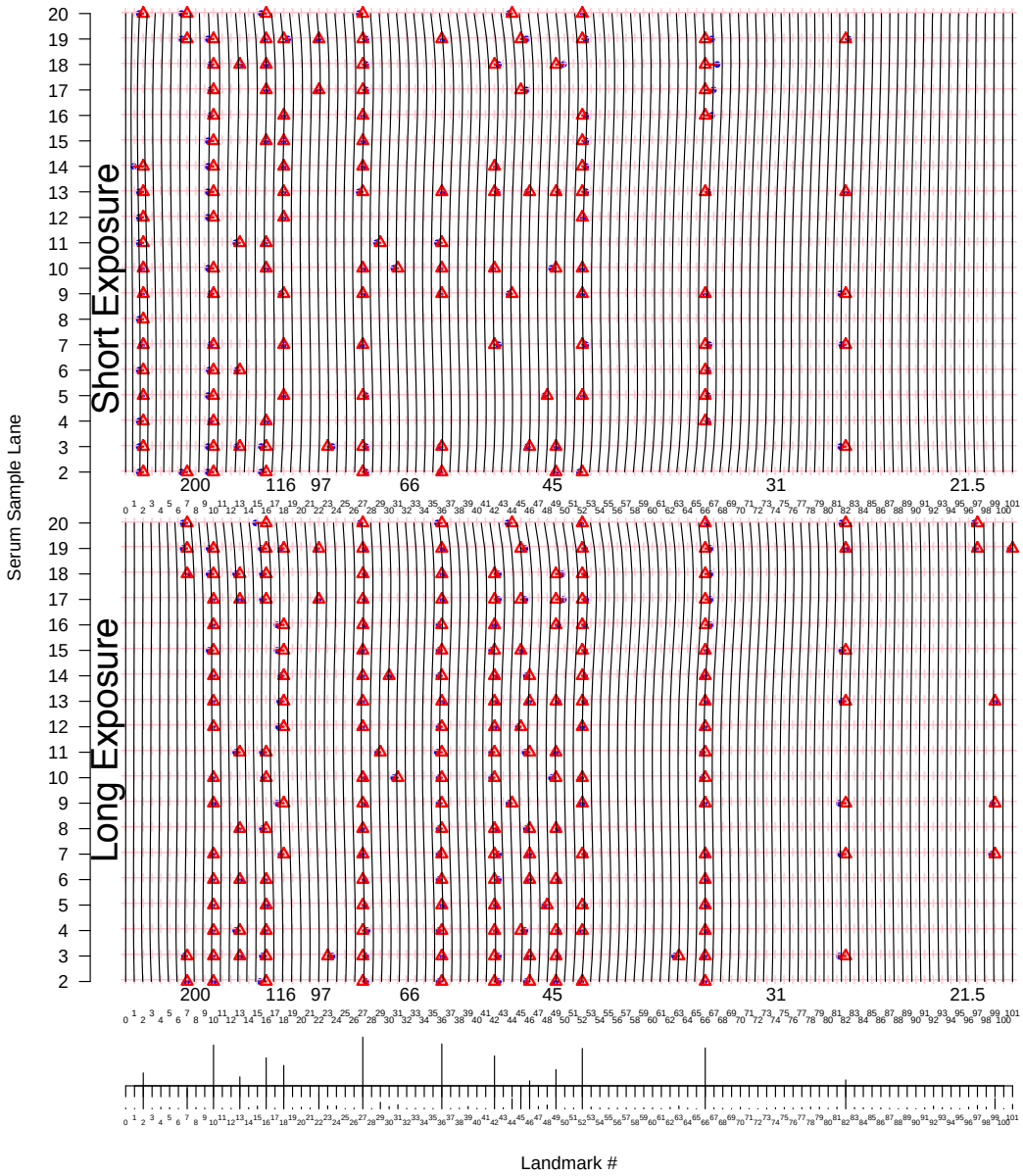

Figure S2. Bayesian gel dewarping for the replication experiment (Top/Middle: short/long exposure; reference lane 1s excluded). *Top:* For each gel set, 19 serum lanes at $L = 100$ interior landmarks. Solid blue dots "•" are detected peaks deviating from its true weight. Each detected peak T_{gij} is connected to a red triangle "Δ" that represents the *maximum a posteriori* molecular weight landmark $\widehat{Z}_{gij}$. The bundle of black vertical curves visualize the deformations, with each black vertical curve connecting estimated locations with identical molecular weights. The curves are drawn for each landmark. *Bottom:* Marginal posterior probabilities of each landmark protein present in a sample.

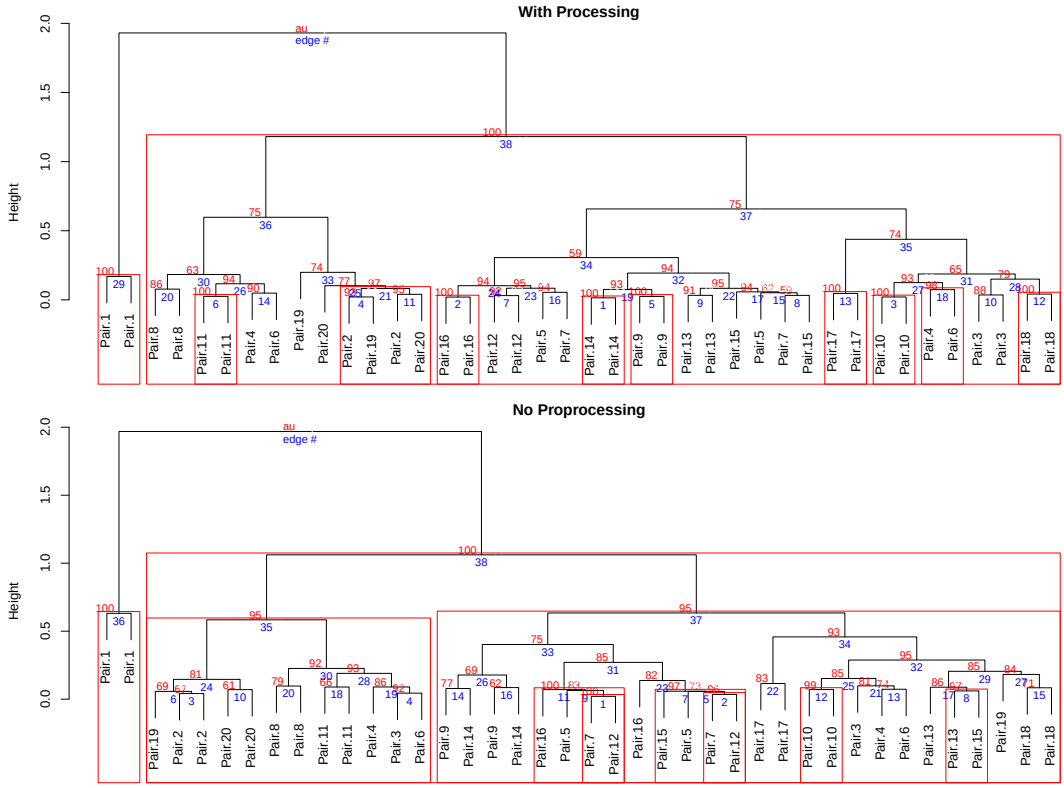

Figure S3. Estimated dendrograms with (top) and without (bottom) preprocessing for the replication experiment. The red boxes show the subtree with $> 95\%$ confidence levels. The actual confidence levels are shown in red on top of the subtrees. The numbers below the edges denote the edge numbers.

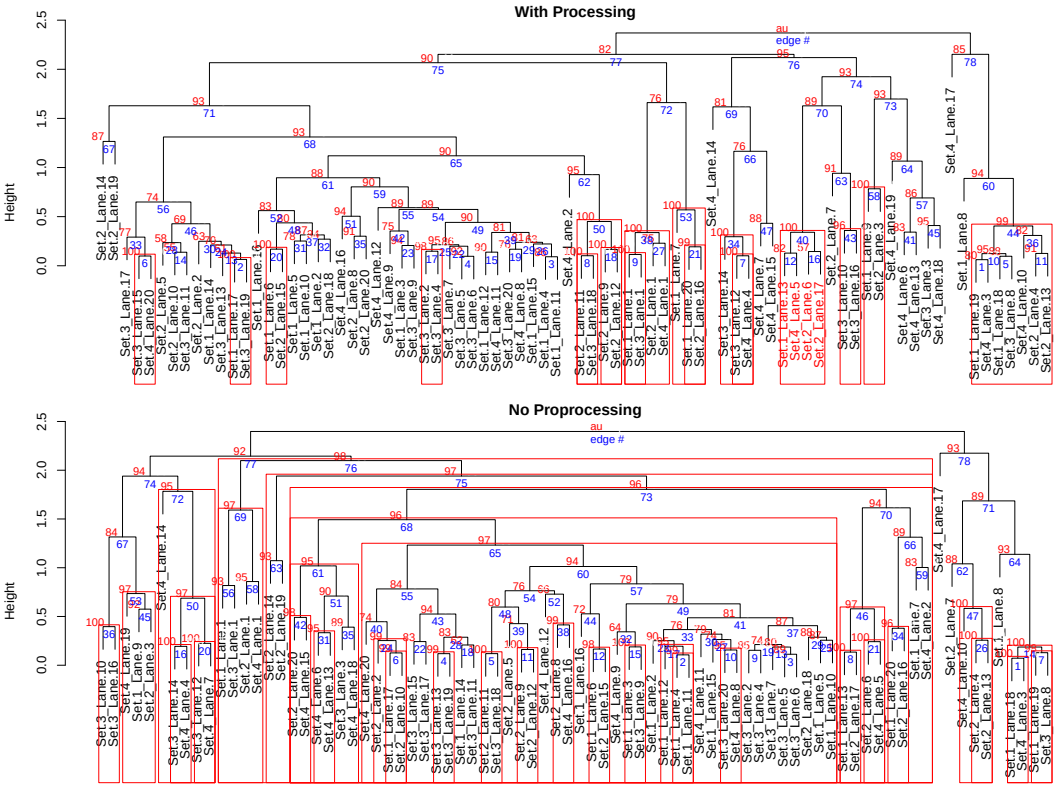

Figure S4. Estimated dendrograms with (top) and without (bottom) preprocessing for the second data set. The red boxes show the subtree appearing in $> 95\%$ of bootstrapped dendrograms with the actual estimated frequencies shown in red on top of the subtrees.

References

Carlson, Nichole E, Grunwald, Gary K and Johnson, Timothy D. (2015). Using cox cluster processes to model latent pulse location patterns in hormone concentration data. *Biostatistics*, kxv046.

Du, Pan, Kibbe, Warren A and Lin, Simon M. (2006). Improved peak detection in mass spectrum by incorporating continuous wavelet transform-based pattern matching. *Bioinformatics* **22**(17), 2059–2065.

Morrissey, Edward R, Juárez, Miguel A, Denby, Katherine J and Burroughs, Nigel J. (2011). Inferring the time-invariant topology of a nonlinear sparse gene regulatory network using fully bayesian spline autoregression. *Biostatistics* **12**(4), 682–694.

Schwartzman, Armin, Gavrilov, Yulia and Adler, Robert J. (2011). Multiple testing of local maxima for detection of peaks in 1d. *Annals of statistics* **39**(6), 3290.

Uchida, Seiichi and Sakoe, Hiroaki. (2001). Piecewise linear two-dimensional warping. *Systems and Computers in Japan* **32**(12), 1–9.