Biostatistics (2017), **0**, 0, *pp.* 1–13
doi:10.1093/biostatistics/gel-supplementary-materials˙rev1

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 $s = 1, \ldots, S_g$, lane $i = 1, \ldots, N_g$, 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 $i = 1, \ldots, N_g$, 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).

*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 $\mathcal{B}^0_{gi}(h) = \{b \mid \mathsf{score}_{gi}(b) = 3\}$. Because h controls the locality of a peak, we perform peak-candidacy search for a few values of h. Denote the union of the identified peak-candidate bins by $\mathcal{B}^0_{gi} = \cup_h \mathcal{B}^0_{gi}(h)$. $\mathcal{B}^0_{gi}$ is comprised of multiple blocks, each comprised 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. We refer to them as *peaks* $j = 1, \ldots, J_{gi}$ 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 peak 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 actin peaks stand out clearly. For example, the middle panel of Figure 1(b) shows the

result of peak detection by blue asterisks for one set of the gels (Section 4.3), where the peaks that are slightly higher than the neighboring intensities 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 $C_0 = 0.01$ that is 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 (e.g., Carlson *and others*, 2015). However, because a minor peak has a narrow span comprised of few bins, its presence or absence is more uncertain than major peaks under random process models. In addition, to detect peaks via the random process models for hundreds of samples and hundreds of dimensions per sample, the iterative MCMC model fitting algorithm could be computationally expensive.

Appendix S3. Reference Alignment via Piecewise Linear Dewarping

Let $(\boldsymbol{t} = (t_1, \ldots, t_B), \boldsymbol{M}_g = \{M_{gib}\})$ represent the high-frequency intensity data for a query gel g and $(\boldsymbol{t}, \boldsymbol{M}_{g_0} = \{M_{g_0ib}\})$ the data for a template gel g_0. We describe a procedure to align gel g towards g_0 using piecewise linear dewarping (Uchida and Sakoe, 2001). Let $\boldsymbol{\kappa}_g$ represent the set of knots comprised of two endpoints $\kappa_{g1} = \nu_0, \kappa_{gR} = \nu_{L+1}$ *and* the reference peaks $\{\kappa_{g2}, \ldots, \kappa_{g,R-1}\}$ on gel g. The knot κ_{gj} corresponds to a bin defined by $\{b_{gj} : t_{b_{gj}} = \kappa_{gj}\}$, $j = 1, \ldots, R$. For the

template gel g_0, we similarly denote the knots and the bins by $\boldsymbol{\kappa}_{g_0}$ and $b_{g_0 j}$, $j = 1, \ldots, R$. Note that $\boldsymbol{\kappa}_g$ and $\boldsymbol{\kappa}_{g_0}$ need to have identical cardinality R. In our application, $R = 9$: seven reference peaks plus two endpoints.

In two steps, we define a piecewise linear function $\mathcal{W}_g(\cdot; g_0)$ for reference alignment. The function first matches the query knots $\boldsymbol{\kappa}_g$ to the template knots $\boldsymbol{\kappa}_{g_0}$ and then linearly stretches or compresses gel g between the matched knots. It finds a bin number $\mathcal{W}_g(b; g_0)$ in the query gel g and match it to bin b in the template gel g_0. For example, we will have $\mathcal{W}_g(b_{g_0 j}; g_0) = b_{gj}$ i.e., the j-th knots in both gels are exactly matched.

For the bin $b = 1, \ldots, B$ on the template gel g_0:

1. Find the nearest knot to its left: $\{j = j(b; g_0) \in \{1, \ldots, R-1\} : \kappa_{g_0 j} \leqslant t_b < \kappa_{g_0, j+1}\}$.

2. Find the bin in the query gel g to be matched by bin b by

$$\mathcal{W}_g^{(j)}(\cdot; g_0) : b \mapsto \lfloor w \cdot b_{g,j+1} + (1 - w) b_{gj} \rfloor, \tag{A1}$$

   where $w = w(b; g_0) = (b - b_{g_0 j}) \,/\, (b_{g_0, j+1} - b_{g_0 j})$, and $\lfloor a \rfloor$ is the largest integer smaller than or equal to a.

We thus obtain the piecewise linear dewarping function

$$\mathcal{W}_g(\cdot; g_0) : b \mapsto \sum_{j=1}^{R-1} \mathcal{W}_g^{(j)}(b; g_0) \mathbf{1}_{\{\kappa_{g_0 j} \leqslant t_b < \kappa_{g_0, j+1}\}}, b = 1, \ldots, B,$$

that aligns the query gel g to the template gel g_0 ($g \neq g_0$).

Figure S1 illustrates the results before and after the piecewise linear dewarping by $\mathcal{W}_g(\cdot; g_0)$. The piecewise linear dewarping automatically matches the reference peaks from multiple gels to facilitate cross-gel band comparisons.

Appendix S4. Details on Shrinkage Prior for Hyperparameters $\{\sigma_{gs}^{-2}\}$

For the hyperpriors on the smoothing parameters $\{\tau_{gs}^2 = \sigma_{gs}^{-2}\}$, $s = 2, \ldots, T_\nu - 1$, $g = 1, \ldots, G$, we specify a two-component mixture distribution with one component favoring small and the other favoring large values (Morrissey *and others*, 2011):

$$\tau_{gs}^2 \sim \xi_{gs}\mathsf{Gamma}(\cdot \mid a_\tau, b_\tau) + (1 - \xi_{gs})\mathsf{InvPareto}(\cdot \mid a'_\tau, b'_\tau), \tag{A2}$$

$$\mathsf{InvPareto}(\tau; a, b) = \frac{a}{b}\left(\frac{\tau}{b}\right)^{a-1}, a > 0, 0 < \tau < b, \tag{A3}$$

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_{gs}^2 so that given s, β_{gst}, $t = 1, \ldots, T_u$ are similar or discrepant according as τ_{gs}^2 being large or small. The random smoothness indicator ξ_s represents different (1) or the same amount of (0) warping across lanes. Let $\xi_{gs} \sim \mathsf{Bernoulli}(\rho_g)$ with success probability ρ_g. We specify a hyperprior $\rho_g \sim \mathsf{Beta}(a_\rho, b_\rho)$ to let data inform the degree of smoothness. In our 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 the submodels defined by $(\xi_{g2}, \ldots, \xi_{g,T_\nu - 1}) \in \{0, 1\}^{T_\nu - 1}$.

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 the actin peaks to a single landmark (*maximum a posteriori*).

Appendix S5. 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

$$\mathbb{P}(Z_{gij} = \ell \mid others) \propto N\left(T_{gij}; \mathcal{S}_g(u_i, \nu_\ell; \boldsymbol{\beta}_g), \sigma_\epsilon^2\right) \{1 - \exp(-\lambda_\ell^*)\}, \tag{A4}$$

for $\ell = 1, \ldots, L$ that satisfy the support constraint $|\nu_\ell - T_{gij}| < A_0$ and order restriction $Z_{gij} < Z_{gi,j+1}$, $j = 1, \ldots, J_{gi} - 1$.

2. Update the B-spline basis coefficients $\boldsymbol{\beta}_g = \{\beta_{gst}\}$ for gel $g = 1, \ldots, G$.

To establish notation, let Δ_1 be the first order difference operator of dimension $(T_\nu - 2) \times T_\nu - 1$ with entries $\Delta_{1kk'} = \delta(k+1, k') - \delta(k, k')$ and $\delta(k, k')$ equals 1 if $k = k'$ and 0 otherwise; Similarly we define Δ_2 with T_ν replaced by $T_u + 1$. The random walk priors (2.6) and (2.7) can then be written as

$$\boldsymbol{\beta}_{g[-T_\nu]1} \overset{d}{\sim} N_{T_\nu - 1}(\cdot; \boldsymbol{\beta}^{\mathsf{id}}_{[-T_\nu]}, \sigma_{g1}^{-2} \Delta_1' \Delta_1) \mathbf{1}\{\nu_0 = \beta_{g11} < \ldots < \beta_{g,T_\nu - 1,1} < \nu_{L+1}\},$$

and

$$\boldsymbol{\beta}_{gs\cdot} \overset{d}{\sim} N_{T_u}(\cdot; \mathbf{0}, \sigma_{gs}^{-2} \Delta_2' \Delta_2) \mathbf{1}\{\nu_0 = \beta_{g1t} < \beta_{g,s-1,t} < \beta_{gst} < \nu_{L+1}, \forall t \geqslant 2\}, s = 2, \ldots, T_\nu - 1.$$

Although both Δ_1 and Δ_2 are rank deficient, we will show that the conditional posterior for $\boldsymbol{\beta}_g$ is proper under a *scattering condition*.

We now write the likelihood (2.2) of the $P_g = \sum_{i=1}^{N_g} J_{gi}$ observed peaks $\boldsymbol{T}_g$. Let $\mathsf{vec}[\boldsymbol{\beta}_g]$ be a vector of length $T_\nu T_u$ that stacks the columns of $\boldsymbol{\beta}_g$. Let $\boldsymbol{B}_g$ be a P_g by $T_\nu T_u$ matrix; the i-th row is defined by the Kronecker product of two sets of B-spline basis functions evaluated at $(\nu_{Z_{gij}}, u_{gi})$: $\boldsymbol{B}_{g2}(u_{gi})' \otimes \boldsymbol{B}_{g1}(\nu_{Z_{gij}})'$.

Update the B-spline basis coefficients $\boldsymbol{\beta}_g = \{\beta_{gst}\}$ for gel $g = 1, \ldots, G$ by the multivariate Gaussian distribution (with constraints (2.4) and (2.5))

$$[\mathsf{vec}[\boldsymbol{\beta}_g] \mid others] \propto \exp\left(-\frac{1}{\sigma_\epsilon^2} \left|\left| \boldsymbol{T}_g - \boldsymbol{B}_g \mathsf{vec}[\boldsymbol{\beta}_g] \right|\right|_2^2\right)$$

$$\times \exp\left(-\frac{1}{\sigma_{g1}^2}\left|\left|\Delta_1^{\mathsf{aug}}\boldsymbol{\beta}^{\mathsf{id,aug}} - \Delta_1^{\mathsf{aug}}\mathsf{vec}[\boldsymbol{\beta}_g]\right|\right|_2^2\right)$$
$$\times \exp\left(-\sum_{s=2}^{T_\nu-1}\frac{1}{\sigma_{gs}^2}\left|\left|\Delta_{2s}^{\mathsf{aug}}\mathsf{vec}[\boldsymbol{\beta}_g]\right|\right|_2^2\right),$$

where $\boldsymbol{\beta}^{\mathsf{id,aug}}$ is a column vector formed by stacking $\boldsymbol{\beta}^{\mathsf{id}}$ for T_u times hence is of length $T_\nu T_u$. Here Δ_1^{aug} is an augmented first-order difference operator applied to a vector of length $T_\nu T_u$. It augments Δ_1 to $[\Delta_1 \mid \mathbf{0}_{(T_\nu-2)\times(T_uT_\nu-T_\nu+1)}]$. Similarly, we define a first-order difference matrix $\Delta_{2s}^{\mathsf{aug}}$ that augments Δ_2 to a $(T_u-1)\times T_uT_\nu$ matrix; we replace the $(s, s+T_\nu, \ldots, s+(T_u-1)T_\nu)$-th columns $\mathbf{0}_{(T_u-1)\times T_\nu T_u}$ by the columns of Δ_2.

The conditional distribution above then simplifies to a multivariate Gaussian distribution with a 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 the precision matrix $\boldsymbol{\Lambda}_g = \sigma_\epsilon^{-2}\boldsymbol{B}_g'\boldsymbol{B}_g + \sigma_{g1}^{-2}\Delta_1^{\mathsf{aug}'}\Delta_1^{\mathsf{aug}} + \sum_{s=2}^{T_\nu-1}\sigma_{gs}^{-2}\Delta_{2s}^{\mathsf{aug}'}\Delta_{2s}^{\mathsf{aug}}$.

The matrix $\boldsymbol{B}_g'\boldsymbol{B}_g$ is of full rank when the observed peaks are well scattered across lanes and along the gels. Let (c_{s_0}, c_{s_0+1}) be the support of the s_0-th B-spline basis $B_{1s_0}(\cdot)$ in the ν-direction. $\boldsymbol{B}_g$ can be rank deficient, for example, when no peak appears in between (c_{s_0}, c_{s_0+1}). That is, we have $\nu_{Z_{gij}} \notin (c_{s_0}, c_{s_0+1})$ and $B_{g1s_0}(\nu_{Z_{gij}}) = 0$, for $i = 1, \ldots, N_g$, $j = 1, \ldots, J_{gi}$. By the definition of $\boldsymbol{B}_g$, it will have constant zeros in the columns s_0, s_0+T_ν,..., and $s_0+(T_\nu-1)T_u$. As a result, the marginal posterior of $\{\beta_{gs_0t}\}_{t=1}^{T_u}$ can only be learned indirectly 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, $\boldsymbol{B}_g$ can be made full rank by reducing the number of basis functions. However, the warping function $\mathcal{S}_g$ will be less flexible. We therefore recommend a minimal number of basis functions necessary to ensure parameter identifiability provided the family of warping functions $\mathcal{S}_g$ is flexible enough to characterize the image deformations (e.g., curvature of the actin peaks). We refer to the condition that $\boldsymbol{B}_g$ being full rank as the

scattering condition.

In our application, failure of the scattering condition is rare. $\mathbf{\Lambda}_g$ 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 $\sum_{s=2}^{T_\nu-1}\sigma_{gs}^{-2}\Delta_{2s}^{\mathsf{aug}'}\Delta_{2s}^{\mathsf{aug}}$ are both sparse square matrices with at most $\mathcal{O}(T_\nu T_u)$ nonzeros, $\mathbf{\Lambda}_g$ preserves the sparsity of $\boldsymbol{B}_g'\boldsymbol{B}_g$. So we use sparse Cholesky factorization of $\mathbf{\Lambda}_g$ to produce its Cholesky factors.

We first block update $\{\beta_{gst}\}_{t=1}^{T_u}$ for $s=2$ from $[\beta_{gst} \mid \beta_{gj_1j_2}, j_1 \neq s, others]$ with the constraint $\beta_{gst} > \beta_{g,s-1,t}$, $t=1,\ldots,T_u$ and continue for $s=3,\ldots,T_\nu-1$. This step requires calculation of inverse of submatrices of $\mathbf{\Lambda}_g^{-1}$ for $T_\nu-2$ times. In our application, computing one such inverse when $T_\nu=10$ and $T_u=6$ requires <0.001 seconds.

3. Update the smoothing parameters $\tau_{gs}^2=\sigma_{gs}^{-2}$ and smoothness selection indicator ξ_{gs} (Appendix S4). First randomly switch ξ_{gs} to ξ_{gs}^* either from 0 to 1, or 1 to 0 for $s=2,\ldots,T_\nu-1$, $g=1,\ldots,G$. For the parameter τ_{gs}^2, we propose its candidate τ_{gs}^{*2} from the log-normal distribution with log-mean parameter τ_{gs}^2. We accept $(\tau_{gs}^{*2},\xi_{gs}^*)$ with probability

$$\alpha_g^{(1)}=\min\left\{1,\frac{p(\boldsymbol{T}_g;\boldsymbol{Z}_g,\boldsymbol{\beta}_g,\tau_{gs}^{*2})\pi(\tau_{gs}^{*2}\mid\xi_{gs}^*)q(\tau_{gs}^2\mid\tau_{gs}^{*2})q(\xi_{gs}\mid\xi_{gs}^*)}{p(\boldsymbol{T}_g;\boldsymbol{Z}_g,\boldsymbol{\beta}_g,\tau_{gs}^2)\pi(\tau_{gs}^2\mid\xi_{gs})q(\tau_{gs}^{*2}\mid\tau_{gs}^2)q(\xi_{gs}^*\mid\xi_{gs})}\right\}, \tag{A5}$$

where $p(\boldsymbol{T}_g;\boldsymbol{Z}_g,\boldsymbol{\beta}_g,\tau_{gs}^{*2})$ denotes the Gaussian likelihood (2.2) and $\pi(\tau_{gs}^2\mid\xi_{gs})$ is the prior distribution (A2).

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

$$\alpha_g^{(2)}=\min\left\{1,\frac{p(\boldsymbol{T}_g;\boldsymbol{Z}_g,\boldsymbol{\beta}_g,\tau_{gs}^{*2})\pi(\tau_{gs}^{*2}\mid\xi_{gs}^*)q(\tau_{gs}^2\mid\tau_{gs}^{*2})}{p(\boldsymbol{T}_g;\boldsymbol{Z}_g,\boldsymbol{\beta}_g,\tau_{gs}^2)\pi(\tau_{gs}^2\mid\xi_{gs})q(\tau_{gs}^{*2}\mid\tau_{gs}^2)}\right\}. \tag{A6}$$

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 $\min\left\{1, \frac{N_{T_\nu-1}\left(\{\beta_{gs1}\}_{s=1}^{T_\nu-1};\mathbf{0},\tau_{g1}^{*2}\mathbf{I}\right)q(\tau_{g1}^2|\tau_{g1}^{*2})}{N_{T_\nu-1}\left(\{\beta_{gs1}\}_{s=1}^{T_\nu-1};\mathbf{0},\tau_{g1}^{2}\mathbf{I}\right)q(\tau_{g1}^{*2}|\tau_{g1}^{2})}\right\}$.

5. Update the smoothness selection hyperparameter ρ_g (Appendix S4), for $g = 1, \ldots, G$ from

$$[\rho_g \mid others] \sim \mathsf{Beta}(a_\rho + \sum_s \mathbf{1}\{\xi_{gs} = 1\}, b_\rho + \sum_s \mathbf{1}\{\xi_{gs} = 0\}). \tag{A7}$$

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%. All the posterior analyses were based on three chains of 10,000 iterations with a burn-in period of 5,000 iterations. We monitor the MCMC convergence via chain histories (visual inspection of good mixing), kernel density plots (unimodal for continuous parameters) and Brooks-Gelman-Rubin statistic $\widehat{R}$ (Brooks and Gelman, 1998) (three chains with random starting values; we used the criterion $\widehat{R} < 1.01$ to declare posterior convergence for every model parameter). Convergence is fast within thousands of burn-in iterations.

Appendix Figures

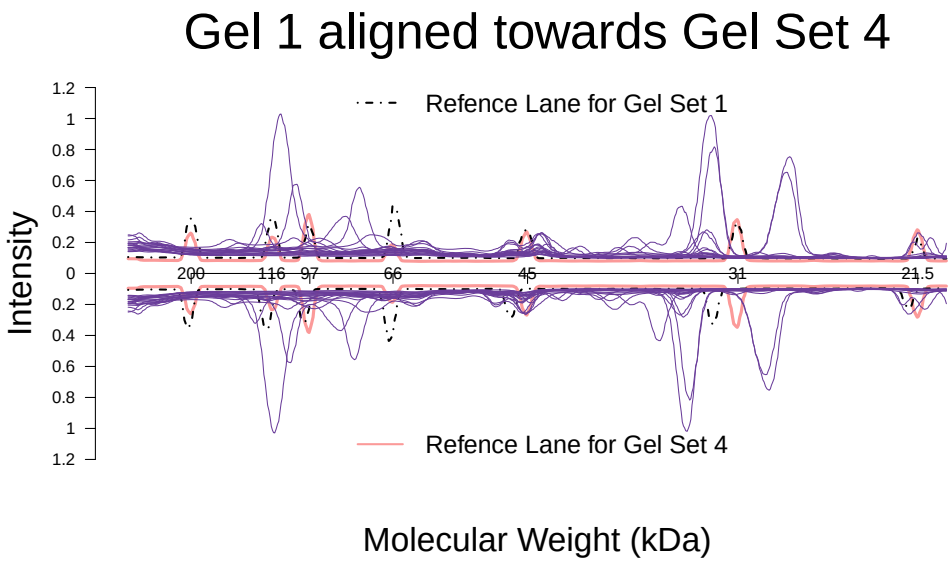

Figure S1. Before (bottom) and after (top) piecewise linear dewarping towards Gel 4 (Section 4.3). *Top*: 21 intensity curves, 20 solid curves from one GEA experiement ($g = 1$) after reference alignment; one dashed curve for the reference lane in gel $g_0 = 4$. The two curves *not* in purple denote 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.

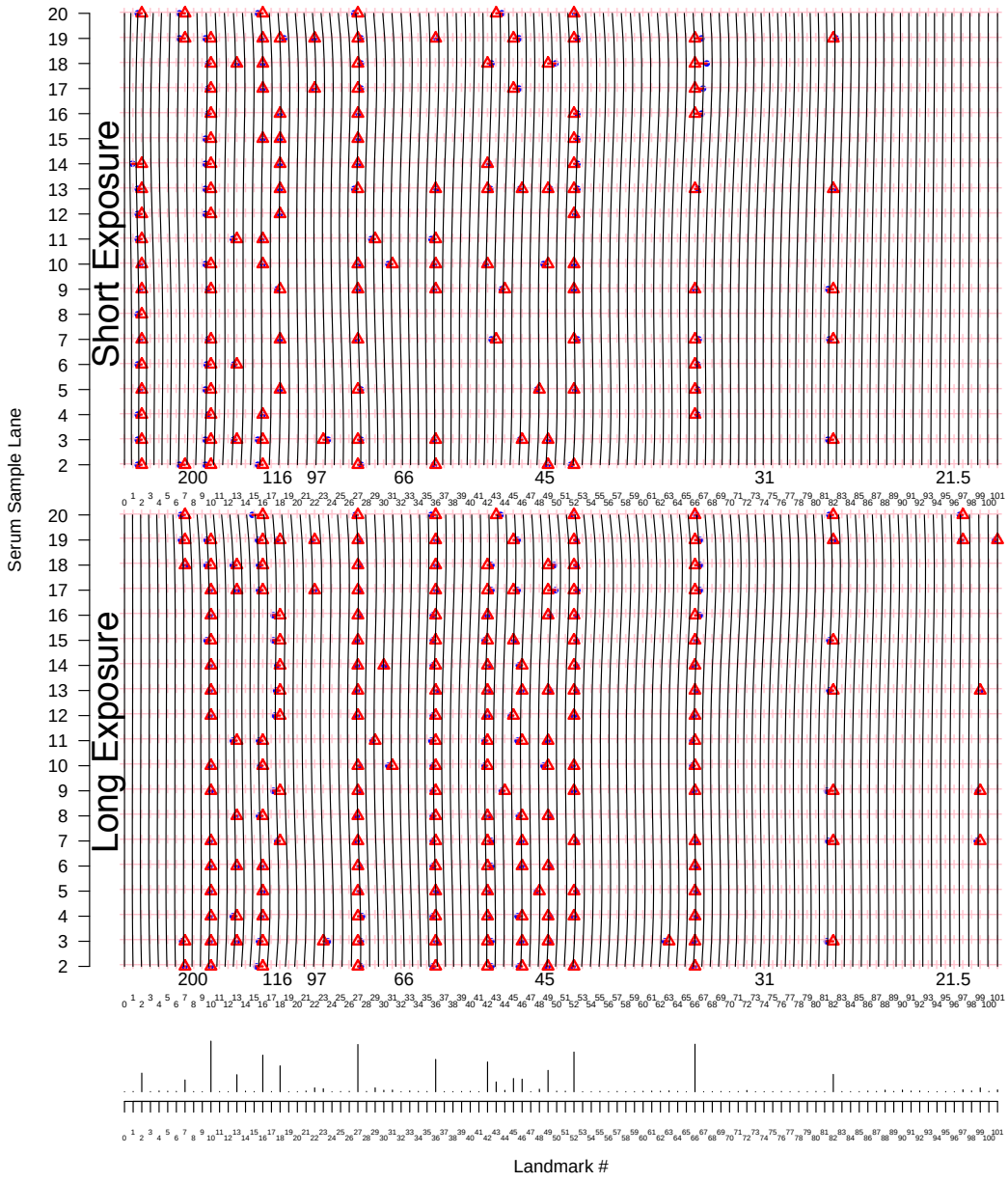

Figure S2. Bayesian spatial dewarping results for the experiment with replicates (short/long exposure). *Top:* For each gel set, 19 serum lanes over a grid of $L = 100$ interior landmarks (reference lanes excluded). Each detected peak T_{gij} (solid blue dots "•") is connected to its *maximum a posteriori* landmark $\widehat{Z}_{gij}$ (red triangle "Δ"). The image deformations are shown by the bundle of black vertical curves $\{u \mapsto \mathcal{S}_g(\nu_\ell, u) : u = 2, \ldots, 20\}$, $\ell = 1, \ldots, L$, $g = 1, 2$, each of which connects the estimated locations of identical molecular weight. *Bottom:* Marginal posterior probability of each landmark protein present in a sample.

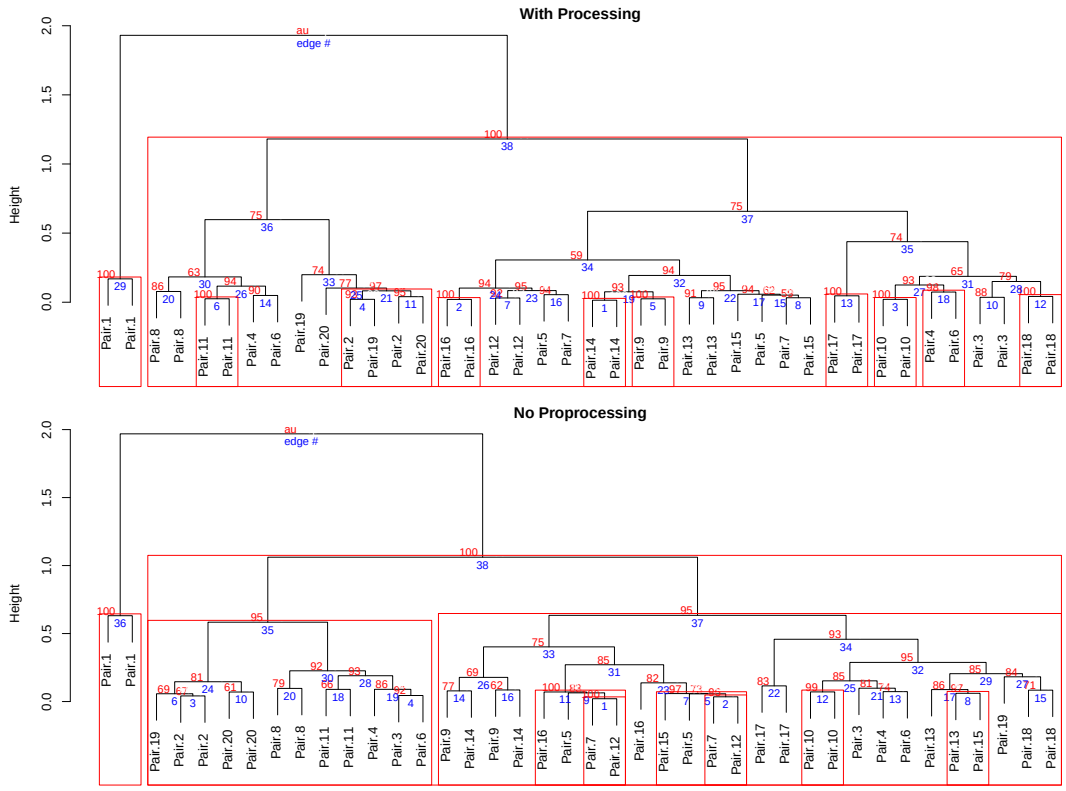

Figure S3. Estimated dendrograms with (top) and without (bottom) pre-processing for the replication experiment. The red boxes show the subtree with $> 95\%$ confidence levels. The actual confidence levels are shown in red on top of the subtrees. The numbers below the edges denote the edge numbers.

References

Brooks, Stephen P and Gelman, Andrew. (1998). General methods for monitoring convergence of iterative simulations. *Journal of computational and graphical statistics* **7**(4), 434–455.

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.