Supplementary Note 2. EM algorithm.

The multivariate model, building on the models introduced in [1, 2], can be expressed as:

\[
p(\hat{\theta}_{UV} \mid R) = N(\hat{\theta}_{UV} \mid \theta_{UV}, \hat{S}_{UV}^{T} R \hat{S}_{UV}^{T})
\]

\[
p(\theta_{g} \mid \Sigma_{1:S}, \pi) = \sum_{m=1}^{M} \sum_{s=1}^{S} \pi_{ms} N(\theta_{g} \mid 0, \omega_{m} \Sigma_{s}).
\]

(1) (2)

We constrain \( \Sigma_{s} \) to factor-model form (see e.g. [3]):

\[
\Sigma_{s} = W_{s} W_{s}^{T} + \Psi_{s},
\]

where \( W_{s} \) is a \( P \times K \) matrix, and \( \Psi_{s} \) a diagonal \( P \times P \) matrix having positive diagonal elements.

Under an improper uniform prior on the factor-model space, our specified prior for \( \Sigma_{s} \) is

\[
p(\Sigma_{s}) \propto \begin{cases} 1 & \Sigma_{s} \in W_{K} \\ 0 & \text{otherwise} \end{cases}
\]

(4)

\[W_{K} := \{ WW^{T} + \Psi_{s} : W \in \mathbb{R}^{P \times K}, \Psi_{s} \text{ diagonal with } [\Psi_{s}]_{jj} \geq 0 \forall j \}\].

(5)

The Dirichlet prior we specify for \( \pi \) is

\[
\vec{\pi} \sim \text{Dirichlet}(\vec{\pi} \mid \alpha_{11}, \ldots, \alpha_{MS}) \quad \alpha_{ms} \geq 1 \forall m, s
\]

(6)

Here we derive an EM algorithm to maximize the log posterior for \((\Sigma_{1:S}, \pi)\) under the model specified at (1)-(6), i.e. to find

\[
\hat{\Sigma}_{1:S}, \hat{\pi} := \arg\max_{\Sigma_{1:S}, \pi} \log p(\Sigma_{1:S}, \pi \mid \hat{\Theta}_{UV}, \hat{R})
\]

or rewritten as the sum of log likelihood and log prior:

\[
\hat{\Sigma}_{1:S}, \hat{\pi} = \arg\max_{\Sigma_{1:S}, \pi} \log p(\hat{\Theta}_{UV} \mid \Sigma_{1:S}, \pi, \hat{R}) + \log[p(\Sigma_{1:S})p(\pi)].
\]

(7)

In the context of the EM algorithm,

\[
D := \delta_{1:P,1:G} \text{ are observed data},
M := \{ \theta_{1:P,1:G}, m_{1:G}, s_{1:G} \} \text{ are latent quantities, and}
P := \{ \Sigma_{1:S}, \pi_{1:M,1:S} \} \text{ are model parameters.}
\]

Note that in \( M \) we have introduced \( m_{1:G} \) and \( s_{1:G} \) as the latent variables defining membership of mixture components, i.e.

\[
\theta_{g} \mid \Sigma_{s}, m_{g} = m, s_{g} = s \sim N(0, \omega_{m} \Sigma_{s})
\]

and that a priori:

\[
P(m_{g} = m, s_{g} = s \mid \pi_{ms}) = \pi_{ms}.
\]

Note also that \( \hat{R} \) is estimated in advance from samples distributed according to the experimental noise process without biological signal; in the case of the International Mouse Phenotyping Consortium (IMPC) data, we randomly sub-sample wild-type animals to create synthetic null data, as described in Methods–Control of error rates.

The optimization at (7), rewritten in the \( D, M, P \) notation above is

\[
\hat{P} = \arg\max_{P} \log p(D \mid P) + \log p(P)
\]

(8)
and the E- and M-steps of the EM algorithm applied iteratively for optimizing (8) are, respectively,

\[ Q(P; P^{(t)}) := E_{M[D, P^{(t)}]} \{ \log p(M, D \mid P) \} \]

\[ P^{(t+1)} = \arg \max_P Q(P; P^{(t)}) + \log p(P) . \]

0.1 E-step

In this stage we evaluate

\[ Q(P; P^{(t)}) := E_{M[D, P^{(t)}]} \{ \log p(M, D \mid P) \} \]

proceeding as follows:

\[
Q(P; P^{(t)}) \\
= \sum_{g} E_{\theta_g, m_g, s_g \mid D_g, P^{(t)}} \log p(\theta_g, m_g, s_g \mid P) \\
= \sum_{g} E_{\theta_g, m_g, s_g \mid D_g, P^{(t)}} \log \prod_{m,s} \left[ p(m_g = m, s_g = s \mid \pi) p(\theta_g, D_g \mid P, m_g = m, s_g = s) \right] \\
= \sum_{g, m,s} E_{\theta_g, m_g, s_g \mid D_g, P^{(t)}} \left[ \log (\pi_{ms} p(\theta_g, D_g \mid P, m_g = m, s_g = s)) \right] \\
= \sum_{g, m,s} E_{m_g, s_g | D_g, P^{(t)}} \left[ \log (\pi_{ms} p(\theta_g, D_g \mid P, m_g = m, s_g = s)) \right] \\
= \sum_{g, m,s} \sum_{m,s} \log (\pi_{ms} + \xi_{gms}) \]

(11)

where

\[ r_{gms}^{(t)} := \frac{\prod_{m_g = m, s_g = s} \{ p(\theta_g, D_g \mid \hat{P}^{(t)}) \}}{\prod_{m_g = m, s_g = s} \{ p(\theta_g, D_g \mid \hat{P}^{(t)}) \}} \]

and

\[ \xi_{gms} := E_{\theta_g | \theta_{sg}^{UV}, m_g = m, s_g = s} \{ \log p(\theta_g, \theta_{sg}^{UV} \mid P, m_g = m, s_g = s) \} . \]

To derive \( \xi_{gms}^{(t)} \) we first note that

\[
\log p(\theta_g, \theta_{sg}^{UV} \mid P, m_g = m, s_g = s) \\
= \log p(\theta_g \mid P, m_g = m, s_g = s) + \log p(\theta_{sg}^{UV} \mid \theta_g, P, m_g = m, s_g = s) \\
= \log N(\theta_g \mid 0, \omega_m \Sigma_s) + \log N(\theta_{sg}^{UV} \mid \theta_g, \hat{S}_{sg}^{UV}) \\
\Sigma_s = \left( \log (\Sigma_s) + \theta_{sg}^{UV} (\omega_m \Sigma_s)^{-1} \theta_g \right) / 2 .
\]

(13)

We then target the distribution for the expectation operator \( E_{\theta_g | \theta_{sg}^{UV}, m_g = m, s_g = s} \) in \( \xi_{gms}^{(t)} \), noting the following joint conditional distribution for \( \theta_g \) and \( \theta_{sg}^{UV} \):

\[
\begin{bmatrix} \theta_g \\ \theta_{sg}^{UV} \end{bmatrix} \mid P^{(t)}, m_g = m, s_g = s \sim N \left( 0, \begin{bmatrix} \omega_m \Sigma_s^{(t)} & \omega_m \Sigma_{s \ast s}^{(t)} \\ \omega_m \Sigma_{s \ast s}^{(t)} & \omega_m \Sigma_{s \ast s}^{(t)} + \hat{S}_{g, s}^{UV} \hat{R}_{s, s}^{UV} \end{bmatrix} \right) 
\]

(14)
so that the conditional distribution \( \theta_g | \theta_{ag}^{UV} \) follows from the conditional multivariate Gaussian identity:

\[
\begin{align*}
\theta_g | \theta_{ag}^{UV}, P^{(t)}, m_g = m, s_g = s & \sim N(\mu_{gm,s}^{(t)}, V_{gm,s}^{(t)}) \\
\mu_{gm,s}^{(t)} & := \omega_m \Sigma_s \cdot (\omega_m \Sigma_s + S_{g,s}^{UV} \hat{R}_{s}^{UV} S_{g,s}^{UV})^{-1} \theta_{ag}^{UV} \\
V_{gm,s}^{(t)} & := \omega_m \Sigma_s - \omega_m \Sigma_s \cdot (\omega_m \Sigma_s + S_{g,s}^{UV} \hat{R}_{s}^{UV} S_{g,s}^{UV})^{-1} \omega_m \Sigma_s.
\end{align*}
\]  

Combining (13) and (15) in (12) gives

\[
\xi_{gms}^{(t)} \triangleq -\frac{1}{2} \text{tr}[\omega_m \Sigma_s | \log |\Sigma_s| + \theta_{ag}^{UV} (\omega_m \Sigma_s)^{-1} \theta_{ag}^{UV}] \\
= -\frac{1}{2} \log |\Sigma_s| - \frac{1}{2} \text{tr}[(\omega_m \Sigma_s)^{-1}(V_{gm,s}^{(t)} + \mu_{gm,s}^{(t)} \mu_{gm,s}^{(t)^T})] .
\]  

Substituting (16) into (11) gives

\[
Q(P; P^{(t)}) = \sum_{g,m,s} r_{gms}^{(t)} \left( \log(\pi_{ms}) - \frac{1}{2} \log |\Sigma_s| - \frac{1}{2} \text{tr}[(\omega_m \Sigma_s)^{-1}(V_{gm,s}^{(t)} + \mu_{gm,s}^{(t)} \mu_{gm,s}^{(t)^T})] \right)
\]  

\[\text{(17)}\]

**0.2 M-step**

In this stage we compute \( \Sigma_s^{(t+1)}, \pi^{(t+1)} \), as per the M-step:

\[
\hat{P}^{(t+1)} = \arg\max_P Q(P; P^{(t)}) + \log p(P) .
\]

The optimisation can performed on \( \Sigma_s^{(t+1)} \) separately for each \( s \); likewise, \( \pi^{(t+1)} \) can be optimised separately.

Addressing first the optimisation with respect to \( \Sigma_s^{(t+1)} \) for one particular \( s \), collecting relevant terms from (17) and from the priors at (4) and (6) gives:

\[
\Sigma_s^{(t+1)} = \arg\max_{\Sigma} \left\{ \text{log } |\Sigma| + \text{tr}(\Sigma^{-1} C_s^{(t)}) : \Sigma \in \mathcal{W}_K \right\}
\]

\[\text{(18)}\]

\[
\mathcal{W}_K := \{ WW^T + \Psi_s : W \in \mathbb{R}^{p \times K}, \Psi_s \text{ diagonal with } |\Psi_s|_{jj} \geq 0 \forall j \}
\]

\[\text{(19)}\]

where

\[
C_s^{(t)} := \frac{\sum_{g,m} r_{gms}^{(t)} (V_{gm,s}^{(t)} + \mu_{gm,s}^{(t)} \mu_{gm,s}^{(t)^T})/\omega_m}{\sum_{g,m} r_{gms}^{(t)}} .
\]

\[\text{(20)}\]

The optimisation at (18)-(19) can be performed using standard factor analysis software (we use \texttt{stats::factanal()} in R).

Finally, to maximize \( Q(P; P^{(t)}) \) with respect to \( \pi \) we find

\[
\pi^{(t+1)} = \arg\max_{\pi \in S} \sum_{m,s} (\tilde{\alpha}_{ms} - 1) \log \pi_{ms}
\]

\[\text{(21)}\]

\[
\tilde{\alpha}_{ms} := \alpha_{ms} + \sum_{g} r_{gms}^{(t)}
\]

\[\text{(22)}\]

\[
S := \left\{ \pi_{1:M,1:S} : \sum_{m,s} \pi_{ms} = 1, \pi_{ms} \geq 0 \right\}
\]

and the objective in (21) has the form of a Dirichlet(\( \tilde{\alpha}_{11}, \ldots, \tilde{\alpha}_{M,S} \)) on \text{vec}(\pi), having the unique mode

\[
\pi_{ms}^{(t+1)} = \frac{\tilde{\alpha}_{ms} - 1}{\sum_{m,s} (\tilde{\alpha}_{ms} - 1)}
\]

\[\text{(23)}\]
provided that \( \tilde{a}_{ms} \geq 1 \ \forall \ m, s \) (and this is implied by the prior on \( \pi \) at (6) satisfying \( \alpha_{ms} \geq 1 \ \forall \ m, s \)). In the case of a uniform prior having \( \alpha_{ms} \equiv 1 \), the update is (with reference to (22) and (23))

\[
\pi_{ms}^{(t+1)} = \frac{\sum_g r_{gms}^{(t)}}{\sum_{g,m,s} r_{gms}^{(t)}}.
\]

(24)

References

