$$P(\mathbf{d},\boldsymbol{\tau}^2|\mathbf{m}) \sim P(\mathbf{m}|\mathbf{d},\boldsymbol{\tau}^2)P(\mathbf{d},\boldsymbol{\tau}^2). \quad (2)$$

The reference effects ε_{rj} are marginalized out in the model, and the choice of the reference sample does not affect the final variance estimates $\boldsymbol{\tau}^2$ (21). Assuming independent observations $\mathbf{m}_j$, given the model parameters, and marginalizing over the ε_{rj}, the likelihood term in Eq. **2** is (21):

$$P(\mathbf{m}|\mathbf{d},\boldsymbol{\tau}^2) = \prod_{tj} \int N(m_{tj}|d_t - \varepsilon_{rj}, \tau_j^2) N(\varepsilon_{rj}|0,\tau_j^2) d\varepsilon_{rj}$$

$$\sim \prod_j (2\pi\tau_j^2)^{-\frac{T}{2}} exp(-\frac{\sum_t (m_{tj}-d_t)^2 - \frac{[\sum_t (m_{tj}-d_t)]^2}{T+1}}{2\tau_j^2}). \quad (3)$$

With non-informative priors for $P(\mathbf{d},\boldsymbol{\tau}^2)$ the posterior of Eq. **2** would reduce to maximum–likelihood-estimation of Eq. **3** as in (21). In this paper, we take full advantage of the prior term to construct the scalable Bayesian online-learning version. Application of the prior forms the basis for sequential updates of the posterior in Eq. **2**. Assuming independent prior terms, a non-informative prior $P(\mathbf{d}) \sim 1$, and inverse Gamma conjugate priors for $\boldsymbol{\tau}^2$ with hyperparameters α_j and β_j (23), the prior takes the form

$$P(\mathbf{d},\boldsymbol{\tau}^2) = P(\mathbf{d})P(\boldsymbol{\tau}^2) \sim \prod_j \Gamma^{-1}(\tau_j^2;\alpha_j,\beta_j). \quad (4)$$

The posterior in Eq. **2** is now fully specified given the likelihood (Eq. **3**), the prior (Eq. **4**), and the probe-specific hyperparameters $\boldsymbol{\alpha} = [\alpha_1,\ldots,\alpha_J]$, $\boldsymbol{\beta} = [\beta_1,\ldots,\beta_J]$.

Our primary interest is in estimating the probe-specific variances $\boldsymbol{\tau}^2$, while $\mathbf{d}$ is a nuisance parameter that could be marginalized out from the model to obtain more robust estimates of $\boldsymbol{\tau}^2$. Since no analytical solution is available and sampling-based marginalization approaches would slow down computation, we obtain a single point estimate for the joint posterior in Eq. **2** as a fast approximation by iteratively optimizing $\mathbf{d}$ and $\boldsymbol{\tau}^2$. A mode for $\mathbf{d}$, given $\boldsymbol{\tau}^2$, is searched for by standard quasi-Newton optimization (24). Then, given $\mathbf{d}$, the variance τ_j^2 follows inverse Gamma distribution with hyperparameters $\hat{\alpha}_j = \alpha_j + \frac{T}{2}$ and $\hat{\beta}_j = \beta_j + \frac{1}{2}(\sum_t (m_{tj}-d_t)^2 - \frac{(\sum_t (m_{tj}-d_t))^2}{T+1}))$. This specifies the prior

$$P(\tau_j^2|\mathbf{m},\mathbf{d}) \sim \Gamma^{-1}(\tau_j^2|\hat{\alpha}_j,\hat{\beta}_j). \quad (5)$$

The point estimate for τ_j^2 is given by the mode at $\tau_j^2 = \hat{\beta}_j/(\hat{\alpha}_j+1)$. The parameters $\mathbf{d}$ and $\boldsymbol{\tau}^2$ are iteratively updated until convergence (< 0.01 change in parameter values in our experiments). The inverse Gamma hyperparameters corresponding to the final $\boldsymbol{\tau}^2$ can be retrieved as $\hat{\alpha}_j = \alpha_j + \frac{T}{2}$ and $\hat{\beta}_j = \tau_j^2(\hat{\alpha}_j + 1)$.

Online-learning of variance hyperparameters The above formulation allows incorporation of prior information of the probes in the analysis and sequential updates where the updated hyperparameters $\hat{\boldsymbol{\alpha}}$, $\hat{\boldsymbol{\beta}}$ from the previous batch provide priors for the next batch through Eq. **2** and the prior in Eq. **4**. In the absence of prior information we shall give equal weight for all probes j at the first batch by setting $\alpha_j = 1$; $\beta_j = 1$ for all j. The final probe-level hyperparameters are obtained by updating $\hat{\boldsymbol{\alpha}}$, $\hat{\boldsymbol{\beta}}$ with new observations at each batch until scanning through the complete data collection.

Affinity estimation The remaining task after learning the probe-specific variances $\boldsymbol{\tau}^2$ is to estimate the probeset-level signal $\mathbf{a}$ and probe affinities $\boldsymbol{\mu}$ in Eq. **1**. Unidentifiability of probe affinities is a well-recognized issue in microarray preprocessing (14), and further assumptions are necessary to formulate an identifiable model. A standard approach, used in the widely used RMA algorithm (14) is to assume that the probes capture the underlying signal correctly on average and the probe affinities sum to zero: $\Sigma_j \mu_j = 0$. We propose a more flexible probabilistic approach where this hard constraint is replaced by soft priors that keep the expected probe affinities at zero but allow higher deviations for the more noisy probes that have a higher variance τ_j^2. To implent this we apply a Gaussian prior $\mu_j \sim N(0,\tau_j^2)$ for the affinities. This allows higher fluctuations for the more noisy probes, which yields a better fit between the probeset-level signal estimate $\mathbf{a}$ and the less noisy probes with smaller τ_j^2. Alternatively, the affinity priors could be determined based on known probe-specific factors, such as GC-content which is a key element in probe affinity estimation in the GC-RMA algorithm (15). As probe performance is affected by a number of factors, however, we prefer the data-driven approach which can accommodate noise from various, potentially unknown sources. This model yields a preliminary estimate for the probeset-level summaries. Based on Eq. **1** we have $a_i = s_{ij} - \mu_j - \varepsilon_{ij} \sim N(s_{ij}, 2\tau_j^2)$. A maximum-likelihood estimate for a_i is obtained as a weighted sum of s_{ij} over the probes j, weighted by the inverse variances: $a_i = \frac{1}{\Sigma_j \frac{1}{2\tau_j^2}} \Sigma_j \frac{1}{2\tau_j^2}(s_{ij})$. The corresponding maximum-likelihood estimate for μ_j at sample i is then given by $\mu_j^{(i)} = s_{ij} - a_i$. Averaging of the affinity estimates across multiple samples yields the maximum-likelihood estimate for the affinities $\boldsymbol{\mu} = [\mu_1,\ldots,\mu_J]$.

Probe summarization The final affinity and variance estimates can be used to summarize the probes according to Eq. **1**. The probeset-level signal a_i is now readily obtained by Eq. **1** as the weighted sum of $s_{ij} - \mu_j$ over the probes j, weighted by the inverse variances: $a_i = \frac{1}{\Sigma_j \frac{1}{\tau_j^2}} \Sigma_j \frac{1}{\tau_j^2}(s_{ij} - \mu_j)$.