A Locally Adaptive Algorithm for Multiple Testing with Network Structure

Ziyi Liang, T. Tony Cai, Wenguang Sun, Yin Xia

Supplementary Material

In this supplement, we discuss additional applications of LASLA in Section S1; the construction of corresponding distance matrices in Section S2; related background review on the sparsity-adaptive weights in Sections S3; implementation details and additional simulation results in Sections S4-S5. Proofs of the main results are collected in Section S6. Finally, Sections S7 and S8 extend the discussion and theoretical analysis to dependent primary statistics.

S1 Additional applications

LASLA has a wide range of applications aside from the network-structured data like the GWAS example discussed in the main article. In this section, we introduce two additional challenging settings: data-sharing regression and integrative inference with multiple auxiliary data sets. In both scenarios, traditional frameworks are not applicable since the auxiliary data $\boldsymbol{U}$ and the primary data $\boldsymbol{T}$ do not match in dimension.

Example 1. Data-sharing high-dimensional regression. Suppose we are interested in identifying genetic variants associated with type II diabetes (T2D). Consider a high-dimensional regression model:

$$\boldsymbol{Y} = \boldsymbol{\mu} + \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \tag{S1.1}$$

where $\boldsymbol{Y} = (Y_1, \dots, Y_n)^\intercal$ are measurements of phenotypes, $\boldsymbol{\mu} = \mu\mathbf{1}^\intercal$ is the intercept, with $\mathbf{1}^\intercal$ being a vector of ones, $\boldsymbol{\beta} = (\beta_1, \dots, \beta_m)^\intercal$ is the vector of regression coefficients, $\boldsymbol{X} \in \mathbb{R}^{n\times m}$ is the matrix of measurements of genomic markers, and $\boldsymbol{\epsilon} = (\epsilon_1, \dots, \epsilon_n)^\intercal$ are random errors.

Both genomics and epidemiological studies have provided evidence that complex diseases may have shared genetic contributions. The power for identifying T2D-associated genes can be enhanced by incorporating data from studies of related diseases such as cardiovascular disease (CVD) and ischaemic stroke. Consider models for other studies:

$$\boldsymbol{Y}^k = \boldsymbol{\mu}^k + \boldsymbol{X}^k\boldsymbol{\beta}^k + \boldsymbol{\epsilon}^k, \tag{S1.2}$$

where the superscript k indicates that the auxiliary data are collected from disease type $k \in [K]$. The notations $\boldsymbol{Y}^k$, $\boldsymbol{\mu}^k$, $\boldsymbol{\beta}^k$, $\boldsymbol{X}^k$ and $\boldsymbol{\epsilon}^k$ have similar explanations as above. The identification of genetic variants associated with T2D can be formulated as a multiple testing problem (2.1), where $\boldsymbol{\theta} = (\theta_i : i \in [m]) = \{\mathbb{I}(\beta_i \neq 0) : i \in [m]\}$ is the primary parameter of interest. The primary and auxiliary data sets are $\boldsymbol{T} = (\boldsymbol{Y}, \boldsymbol{X})$ and $\boldsymbol{U} = \{(\boldsymbol{Y}^k, \boldsymbol{X}^k) : k \in [K]\}$, respectively. The auxiliary data $\boldsymbol{U}$ can provide useful guidance by prioritizing the shared risk factors and genetic variants.

Example 2. Integrative "omics" analysis with multiple auxiliary data sets. The rapidly growing field of integrative genomics calls for new frameworks for combining various data types to identify novel patterns and gain new insights. Related examples include (a) the analysis of multiple genomic platform (MGP) data, which consist of several data types, such as DNA copy number, gene expression, and DNA methylation, in the same set of the specimen (Cai et al., 2016); (b) the integrated copy number variation (iCNV) caller that aims to boost statistical accuracy by integrating data from multiple platforms such as whole exome sequencing (WES), whole genome sequencing (WGS) and SNP arrays (Zhou et al., 2018); (c) the integrative analysis of transcriptomics, proteomics and genomic data (Medina et al., 2010). The identification of significant genetic factors can be formulated as (2.1) with mixed types of auxiliary data.

S2 Forming local neighborhoods: illustrations

Recall that, in Section 1, LASLA first summarizes the structural knowledge in a distance matrix $\boldsymbol{D} \in \mathbb{R}^{m\times m}$ where m is the number of hypotheses. The distance matrix describes the relation between each pair of hypotheses in the light of the auxiliary data. For the GWAS example detailed in Section 1, $\boldsymbol{D} = (1 - r_{ij}^2 : i, j \in [m])$ where r_{ij} measures the linkage disequilibrium between the two SNPs i and j.

In Example 1 (data-sharing regression) from Section S1, we can extract the structural knowledge provided by the related regression problems via Mahalanobis distance (Krusińska, 1987). Specifically, let $\{\hat{\boldsymbol{\beta}}^k = (\hat{\beta}_1^k, \dots, \hat{\beta}_m^k)^\intercal : k \in [K]\}$ denote the estimation of $\{\boldsymbol{\beta}^k = (\beta_1^k, \dots, \beta_m^k)^\intercal : k \in [K]\}$. Denote by $\hat{\boldsymbol{\beta}}_i = (\hat{\beta}_i^k : k \in [K])$ the vector of estimated coefficients for the ith genomic marker across K different studies. The distance matrix $\boldsymbol{D} = (D_{ij})_{i,j\in[m]}$ is then constructed via Mahalanobis distance with $D_{ij} = (\hat{\boldsymbol{\beta}}_i - \hat{\boldsymbol{\beta}}_j)\hat{\Sigma}_{\boldsymbol{\beta}}^{-1}(\hat{\boldsymbol{\beta}}_i - \hat{\boldsymbol{\beta}}_j)^\intercal$, where $\hat{\Sigma}_{\boldsymbol{\beta}}$ is the estimated covariance matrix based on $\{\hat{\boldsymbol{\beta}}_i : i \in [m]\}$. Similarly, in Example 2 (analysis with multiple auxiliary data sets), suppose we collect a multivariate variable $\boldsymbol{U}_i$ from different platforms as the side information for gene i, then the Mahalanobis distance can be used to construct a distance matrix $\boldsymbol{D} = (D_{ij})_{i,j\in[m]}$ with $D_{ij} = (\boldsymbol{U}_i - \boldsymbol{U}_j)\hat{\Sigma}_U^{-1}(\boldsymbol{U}_i - \boldsymbol{U}_j)^\intercal$, where $\hat{\Sigma}_U$ is the estimated covariance matrix based on the auxiliary sample $\{\boldsymbol{U}_i : i \in [m]\}$.

We emphasize that LASLA is not limited to the aforementioned examples. Most of the traditional covariate-assisted methods focus on the array-like auxiliary data $\boldsymbol{U} = \{U_i : i = 1, 2, \ldots\}$ that matches primary data coordinate by coordinate. LASLA can also handle this dimension-matching side information as the latter can be represented by a distance matrix $\boldsymbol{D}$ through simple manipulations. Below, we provide a list of practical types of side information and their corresponding methods for constructing the distance matrix.

(a) A vector of categorical covariates. The elements in $\boldsymbol{U}$ take discrete values and the local neighborhoods can be defined as groups. With suitably chosen weights LASLA reduces to the methods considered in Hu et al. (2010), Li and Barber (2019), and Xia et al. (2020) that are developed for multiple testing with groups.

(b) A vector of continuous covariates. We can define distance as either the absolute difference or the standardized difference in rank $D_{ij} = |\hat{F}_m(U_i) - \hat{F}_m(U_j)|$, where $\hat{F}_m(t)$ is the empirical CDF.

(c) Spatial locations. Such structures have been considered in, for example, Lynch et al. (2017), Lei and Fithian (2018) and Cai et al. (2022). The locations are viewed as covariates and D_{ij} is the Euclidean distance between locations i and j.

(d) The correlations in a network or partial correlations in graphical models. See the GWAS example discussed in Section 1 of the main article.

(e) Multiple auxiliary samples. The Mahalanobis distance or its generalizations (Krusińska, 1987) can be used to calculate the distance matrix $\boldsymbol{D}$.

Note that in practical applications, it could be beneficial to "standardize" the distance matrix $\boldsymbol{D}$; this step ensures algorithm robustness. A more comprehensive discussion on the implementation details is relegated to Section S4.1.

S3 Details on sparsity-adaptive weights

Recall the definition from Section 2.2 that the primary statistics T_i has the hypothetical mixture distribution:

$$F_i^*(t) = (1 - \pi_i^*)F_0(t) + \pi_i^* F_{1i}^*(t)$$

for $i \in [m]$. The quantity π_i^* indicates the sparsity level of signals at location i, and π_i^* is allowed to be heterogeneous across m testing locations.

The key idea in existing weighted FDR procedures such as GBH (Hu et al., 2010), SABHA (Li and Barber, 2019) and LAWS (Cai et al., 2022) is to construct weights that leverage π_i^* by prioritizing the rejection of the null hypotheses in groups or at locations where signals appear to be more frequent. Specifically, SABHA defines the weight as $w_i^{\text{sabha}} = 1/(1 - \pi_i^*)$, and LAWS as $w_i^{\text{laws}} = \pi_i^*/(1 - \pi_i^*)$. The sparsity adaptive-weights have an intuitive interpretation. Consider the LAWS weight w_i^{laws}, if π_i^* is large, indicating a higher occurrence of signals at location i, the weighted p-value $P_i^w := P_i/w_i^{\text{laws}} = (1 - \pi_i^*)P_i/\pi_i^*$ will be smaller, up-weighting the significance level of hypothesis i. However, compared to the proposed weights, such weighting scheme ignores structural information in alternative distributions as discussed in Section 2.5.

S4 Additional numerical results with marginally independent data

In this section, we provide the numerical implementation details and collect additional simulation results for data-sharing high-dimensional regression, latent variable model and multiple auxiliary samples under the marginal independence assumption (A1).

S4.1 Implementation Details

In all of our numerical results, the bandwidth h for the kernel estimations in (2.5) and (2.6) is chosen automatically by applying the `density` function with the option "SJ-ste" in R package `stats`. For the size of neighborhoods $m^{1-\epsilon}$, the default choice for ϵ is 0.1 for marginally independent p-values, while for dependent p-values, we set $\epsilon = 0$ to comply with our FDR control theory under weak dependence in Section S7. For the screening parameter τ, we choose the threshold through the BH algorithm with FDR level $\alpha = 0.8$. This choice ensures that the screening set $\{i \in [m] : P_i > \tau\}$ is predominantly composed of null indices. See Cai et al. (2022) for a more comprehensive discussion on the choice of τ.

To enhance algorithmic robustness and numerical performance, we perform a data-driven scaling of the distance matrix $\boldsymbol{D}$ by a constant factor a. A practical guideline is to ensure that the spread of entries in the scaled distance matrix $\boldsymbol{D}/a$ is similar to that of the entries in $\boldsymbol{T}$. We use the interquartile range (IQR) to measure data spread, a strategy similarly employed in Scott (1992). All additional details can be found in the public repository containing all experiments implementation at `https://github.com/ZiyiLiang/r-Blasla`.

S4.2 Heterogeneous alternative distributions

As highlighted in Section 2.5, LASLA can handle hypothesis-specific alternative densities, unlike many popular methods that only accommodate heterogeneous sparsity levels (Li and Barber, 2019; Cai et al., 2022). Building on the discussion in Section 2.5, we present two additional examples involving heterogeneous data to further demonstrate LASLA's strength in leveraging hypothesis-specific information. Consistent with the analysis in Section 2.5, we compare the oracle LASLA procedure with the oracle LAWS to illustrate the methodological distinctions without the influence of practical implementation.

Extending the analysis of Example 1 in Section 2.5, which considers asymmetry in the sign of the signal, we explore a more challenging setting that allows for varying levels of asymmetry across hypotheses.

Example S4.1. Set $F^*_{1i}(t) = \gamma_i N(3,1) + (1-\gamma_i)N(-3,1)$, where γ_i controls the relative proportions of positive and negative signals for the ith hypothesis. Each γ_i follows a uniform distribution over interval $[0.5-r, 0.5+r]$, i.e., $\gamma_i \sim U(0.5-r, 0.5+r)$. We vary r from 0 to 0.5 by 0.1.

In practice, the heterogeneity of alternative densities can be complex, potentially involving a mixture of factors such as signal strength, signal sign, and the shape of the alternative density. In the following setting, we introduce substantial heterogeneity across each of these components:

Example S4.2. Set $F^*_{1i}(t) = \gamma_i N(\mu_i, \sigma_i^2) + (1 - \gamma_i)N(-\mu_i, \sigma_i^2)$, where each parameter follows a uniform distribution: $\gamma_i \sim U(0, 1)$; $\mu_i \sim U(2.5, 3.5)$ and $\sigma_i \sim \mathrm{U}(0.1, 1)$.

The results for the two examples above are summarized in Figure S4.2 and Figure S4.2, respectively. We again observe LASLA's ability in capturing the heterogeneities, thereby improving the power across the settings.

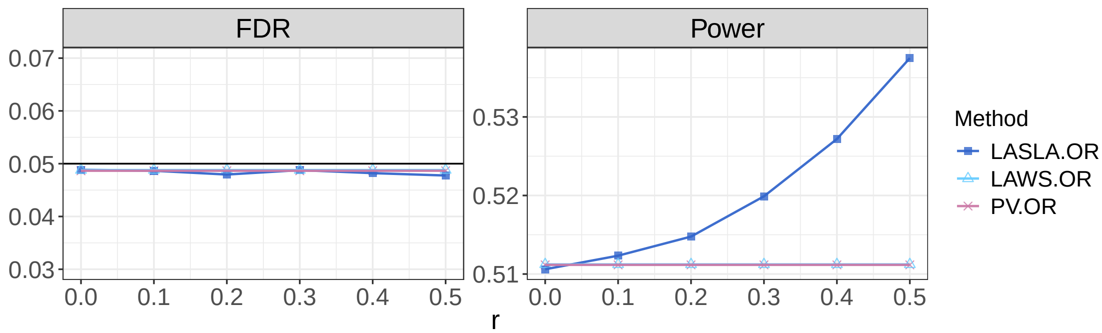

Figure S4.1: Comparison of oracle LASLA and benchmarks with nominal FDR level $\alpha = 0.05$ under the heterogeneous asymmetry setting in Example S4.1. Parameter r controls the level of asymmetry in each individual hypothesis. All other details remain consistent with Figure 2.

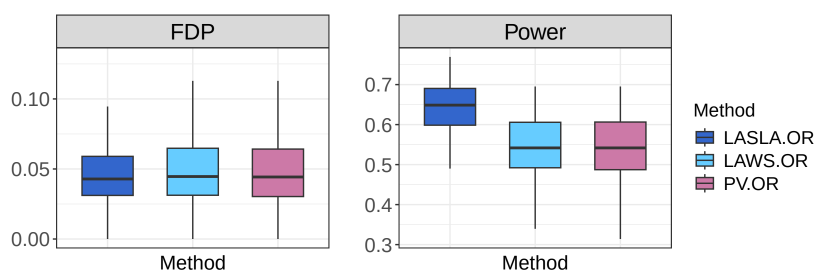

Figure S4.2: Comparison of oracle LASLA and benchmarks with nominal FDR level $\alpha = 0.05$ under the setting in Example S4.2, which introduces substantial heterogeneity across hypotheses.

S4.3 Data-sharing high-dimensional regression

Example 2 in Section S1 discussed how the knowledge in regression models from related studies can be transferred to improve the inference on regression coefficients from the primary model. This section designs simulation studies to illustrate the point.

Consider the regression model (S1.1) defined in Section S1 with $\boldsymbol{X}_{ij} \sim N(0,1)$ for $i \in [n], j \in [m]$ where $\boldsymbol{X}_{ij}$ denotes the entry of $\boldsymbol{X}$ at coordinate (i,j); $\epsilon_i \sim N(0,1)$ for $i \in [n]$. Let $\mathbb{P}(\beta_i = 0) = 0.9$. For the non-null locations, $\beta_i \sim (-1)^u |N(\mu, 0.1)|$; $u \sim \text{Bernoulli}(0.2)$. Note that signals will be more likely to take positive signs, hence asymmetric rejection rules are desired.

Models from K related studies are generated by (S1.2). If the auxiliary model is closely related to the primary model, they tend to share similar coefficients. Therefore, we generate the coefficients for study $k \in [K]$ as $\boldsymbol{\beta}^k = \boldsymbol{\beta} + \boldsymbol{\sigma}$, where each coordinate of $\boldsymbol{\sigma}$ is drawn from normal distribution $N(0, \sigma^2)$. Other quantities are defined similarly as the primary model.

We compute the distance matrix $\boldsymbol{D}$ using the Mahalanobis distance on the estimated coefficients as specified in Section S2. Fix $K = 3, n = 1000, m = 800$, consider the following settings:

- Setting 1: Fix $\mu = 0.25$, vary the noise level σ from 0.1 to 0.2 by 0.02.
- Setting 2: Fix $\sigma = 0.15$, vary the signal strength μ from 0.25 to 0.3 by 0.01.

We compare data-driven LASLA with BH and AdaPT method (Yurko et al., 2020). To apply LASLA, it's essential to have knowledge of the null distribution for the test statistics. In this simulation we use the ordinary least square estimators and T_i follows a t-distribution. Alternatively, one can explore the approach outlined in Xia et al. (2020), where the test statistics follow the

$N(0,1)$ distribution asymptotically. Figure S4.3 shows that LASLA can effectively leverage the side information from related studies and outperforms both BH and AdaPT.

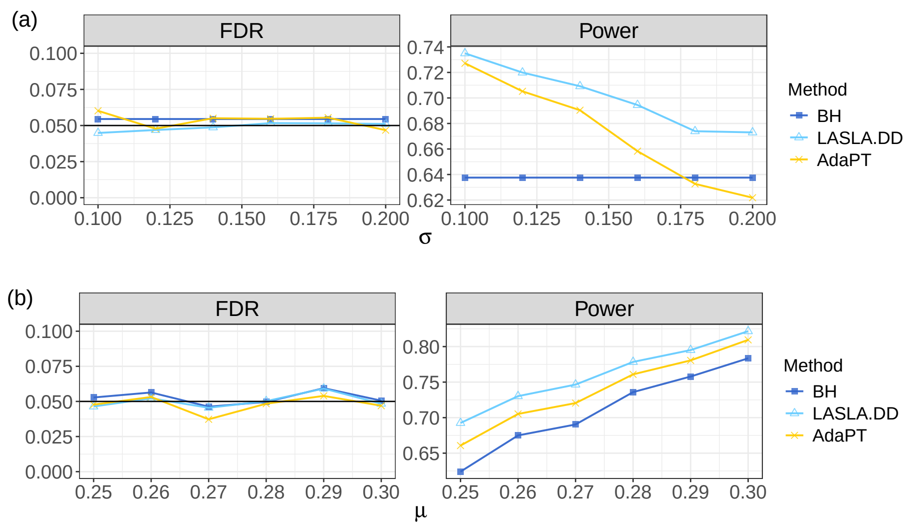

Figure S4.3: Empirical FDR and power comparison of data-driven LASLA and benchmarks at nominal FDR level $\alpha = 0.05$ under the data-sharing regression setting in Section S4.3. (a): Regression setting 1: increasing the noise level σ in the auxiliary data; (b): Regression setting 2: increasing signal strength μ.

S4.4 Latent variable setting

Suppose the primary and auxiliary data are associated with a common latent variable $\boldsymbol{\xi} = (\xi_i : i \in [m])$ where $\xi_i \sim (1-\theta_i)\Delta_0 + \theta_i N(\mu, 1)$ and Δ_0 is the Dirac delta function, namely, $\xi_i = 0$ if $\theta_i = 0$. The primary data $\boldsymbol{T} = (T_i : i \in [m])$ and auxiliary data $\boldsymbol{U} = (U_i : i \in [m])$ respectively follow:

$$T_i \sim N(\xi_i, 1), \quad U_i \sim N(\xi_i, \sigma_s^2), \tag{S4.3}$$

where σ_s controls the informativeness the auxiliary data. Our goal is to test m hypotheses on θ_i as stated in (2.1). Fix $m = 1200$ and let $\theta_i \overset{\text{ind}}{\sim} \text{Bernoulli}(0.1)$, for $i \in [m]$. We consider two settings:

- Setting 1: Fix $\mu = 2.5$, vary σ_s from 0.5 to 2 by 0.25.

- Setting 2: Fix $\sigma_s = 1$, vary μ from 3 to 4 by 0.2.

We compute the distance matrix $\boldsymbol{D}$ from the auxiliary data $\boldsymbol{U}$ using the Euclidean distance, i.e. $D_{ij} = |U_i - U_j|$. We then compare LASLA with the BH procedure, data-driven SABHA (SABHA.DD) as reviewed in Section S3 and AdaPT method (Yurko et al., 2020).

The results are summarized in Figure S4.4. In both settings, LASLA achieves a lower FDR than SABHA while still outperforming it in power. This is because SABHA relies solely on p-values and uses weights that only account for sparsity. The AdaPT method performs comparably to data-driven LASLA when the noise level in the auxiliary data is low, but its performance rapidly deteriorates as the auxiliary information becomes noisier.

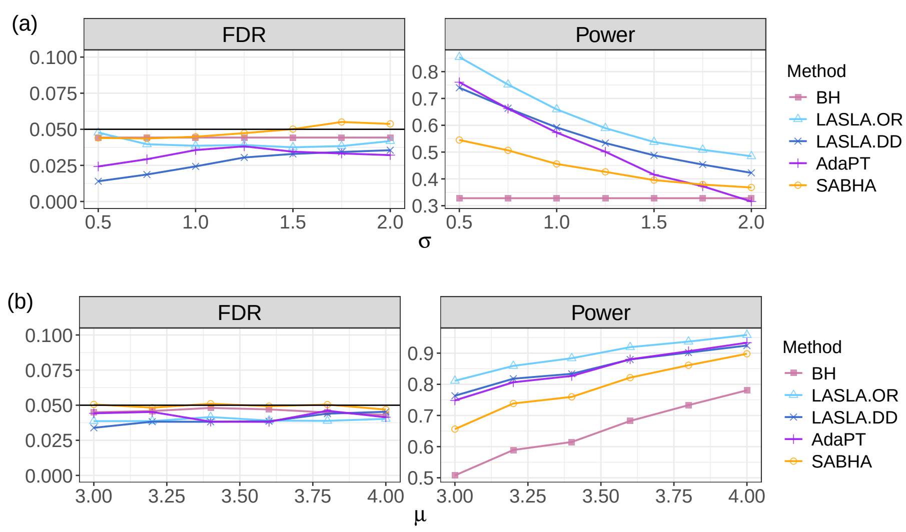

Figure S4.4: Empirical FDR and power comparison of LASLA and benchmarks at nominal FDR level $\alpha = 0.05$ under the latent variable setting in Section S4.4. (a): Latent setting 1: increasing the noise level σ in the auxiliary data; (b): Latent setting 2: increasing signal strength μ.

S4.5 Multiple auxiliary samples

We explore two scenarios with multiple auxiliary samples: (1) all samples are informative; (2) some samples are non-informative. Similar to the previous section, consider a latent variable $\boldsymbol{\xi} = (\xi_i : i \in [m])$ where $\xi_i \sim (1-\theta_i)\Delta_0 + \theta_i N(\mu, 1)$ and $\theta_i \sim \text{Bernoulli}(0.1)$. The primary statistics $T_i \sim N(\xi_i, 1)$ for $i \in [m]$. The goal is to make inference on the unknown θ_i. Let $\boldsymbol{U}^k = (U_i^k : i \in [m])$ denote the kth auxiliary sequence for $k \in [K]$. If $\boldsymbol{U}$ is informative, it should carry knowledge on the underlying signal θ_i. Hence we introduce the first setting where all auxiliary samples are associated with the latent variable $\boldsymbol{\xi}$:

- Setting 1: $U_i^k \sim N(\xi_i, \sigma_s^2)$ for $i \in [m], k = 1, ..., 4$.

Let $\gamma_i \sim \text{Bernoulli}(0.1)$ for $i \in [m]$ independently of everything else, and $\psi_i \sim (1-\gamma_i)\Delta_0 + \gamma_i N(\mu, 1)$. Consider:

- Setting 2: $U_i^k \sim N(\xi_i, \sigma_s^2)$, for $k = 1, 2$; $U_i^k \sim N(\psi_i, \sigma_s^2)$, for $k = 3, 4$.

Note that γ_i being independent of θ_i can lead to significant divergence between the latent variables ψ_i and ξ_i, potentially making $\boldsymbol{U}^3$ and $\boldsymbol{U}^4$ anti-informative. The construction of $\boldsymbol{D}$ from $\{\boldsymbol{U}^k\}_{k\in[K]}$ is not unique, we explore two different methods: using Mahalanobis distance vs using Euclidean distance with the averaged data $U_i^{\text{avg}} = \frac{1}{4}(U_i^1 + U_i^2 + U_i^3 + U_i^4)$ for $i \in [m]$. We assess their effectiveness under varying degrees of informativeness exhibited by the auxiliary samples.

In both settings, we fix $m = 1200, \mu = 3$, and change σ_s from 0.5 to 2 by 0.25. The results are summarized in Figure S4.5. Intuitively, the averaging method reduces variance when all auxiliary samples are informative and leads to power gain over the Mahalanobis approach. However, the latter appears to be more robust when some samples are anti-informative.

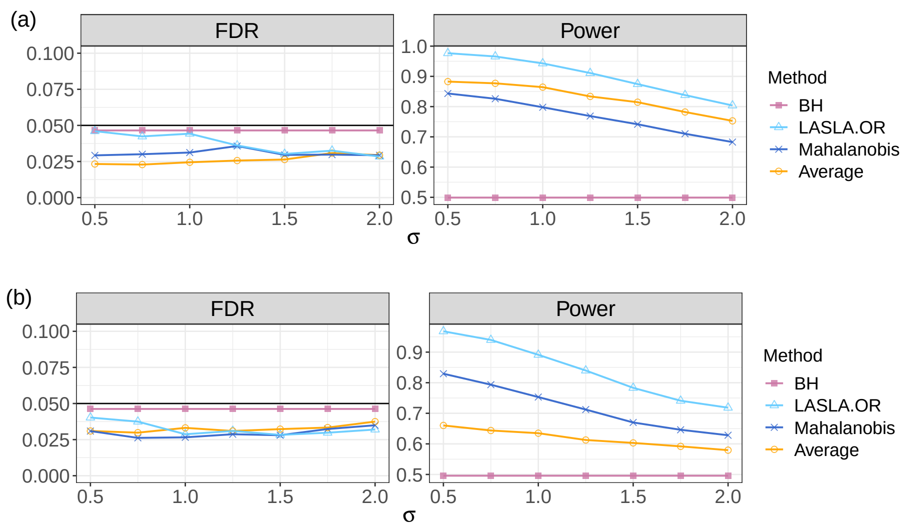

Figure S4.5: Empirical FDR and power comparison of LASLA with different distance computations at a nominal FDR level of $\alpha = 0.05$ under the multiple auxiliary sample setting in Section S4.5. (a): Multiple auxiliary samples, setting 1: all auxiliary samples are informative; (b): Multiple auxiliary samples, setting 2: half of the auxiliary samples are uninformative.

S4.6 Comparison of alternative thresholding rules

As mentioned in Remark 1 of Section 2.4, it is possible to substitute the LASLA thresholding rule in (2.11) with other types of weighted thresholding rules such as the weighted BH procedure (WBH) (Genovese et al., 2006) and the adaptively adjusted WBH procedure (Adj-WBH) (Ramdas et al., 2019). In this section, we first describe both approaches and then present numerical comparisons.

The WBH method applies the BH procedure to the weighted p-values $\{P_i^w = P_i/w_i : i \in [m]\}$ at level α, while the Adjusted WBH method (Ramdas et al., 2019) improves on WBH by applying the BH procedure at an adaptively adjusted FDR level. Specifically, it computes a "weighted sparsity estimator"

$$\hat{\pi} = 1 - \frac{|\mathbf{w}|_\infty + \sum_{i=1}^m w_i I\{P_i > \tau\}}{m(1-\tau)} \tag{S4.4}$$

where $\mathbf{w} = (w_1, \ldots, w_m)$ are the weights, $|\cdot|_\infty$ is the infinity norm and $\tau > 0$ is a screening parameter similar to the one used in (2.5). Note that the weighted sparsity estimator does not account for potential heterogeneity in the sparsity level across the hypotheses. When such heterogeneity is significant, the adaptive adjustment may not be optimal, as we shall see later in Figure S4.7. The rejection threshold of the adjusted WBH method is then computed by applying the BH procedure to the weighted p-values $\{P_i^w = P_i/w_i : i \in [m]\}$ at level $\alpha/(1-\hat{\pi})$.

Next, we provide some numerical comparisons of different thresholding approaches. Under the latent variable setting detailed in Section S4.4, we compare LASLA with WBH and adjusted WBH, both benchmarks are implemented with data-driven LASLA weights. The purpose of this comparison is to isolate the effects of different thresholding rules, as the weights remain consistent across methods. Figure S4.6 shows that WBH is overly conservative, whereas both LASLA and adjusted WBH achieve higher power by adjusting to the sparsity level in the dataset. We emphasize that in this latent variable setting, the sparsity level is constant across all locations $i \in [m]$, hence LASLA and adjusted WBH have nearly identical performance. In the next experiment, we examine a synthetic setting with heterogeneous sparsity levels.

Consider a simple scenario where the primary data $\boldsymbol{T} = (T_i : i \in [m])$ are generated as $T_i \sim (1-\theta_i)N(0,1) + \theta_i N(2,1)$, where $\theta_i \sim \text{Bernoulli}(\pi_i)$ with

$$\pi_i \sim U(0.8, 0.9), \text{ for } i = 1, \ldots, 200; \quad \pi_i = 0.01, \text{ for } i = 201, \ldots, 1000. \tag{S4.5}$$

Here, the sparsity levels are heterogeneous, with elevated levels in the first 200 indices. For simplicity and direct methodological comparison, we use oracle quantities across all methods in this heterogeneous setting. Both adjusted and unadjusted WBH methods utilize the oracle LASLA weights, and the weighted non-null proportion in (S4.4) is also computed using the oracle LASLA weights. The results in Figure S4.7 demonstrate that LASLA effectively utilizes the heterogeneity in sparsity levels and outperforms the adjusted WBH method. The adjusted WBH falls short because the weighted sparsity level in (S4.4) only captures the global sparsity level, neglecting potential heterogeneity.

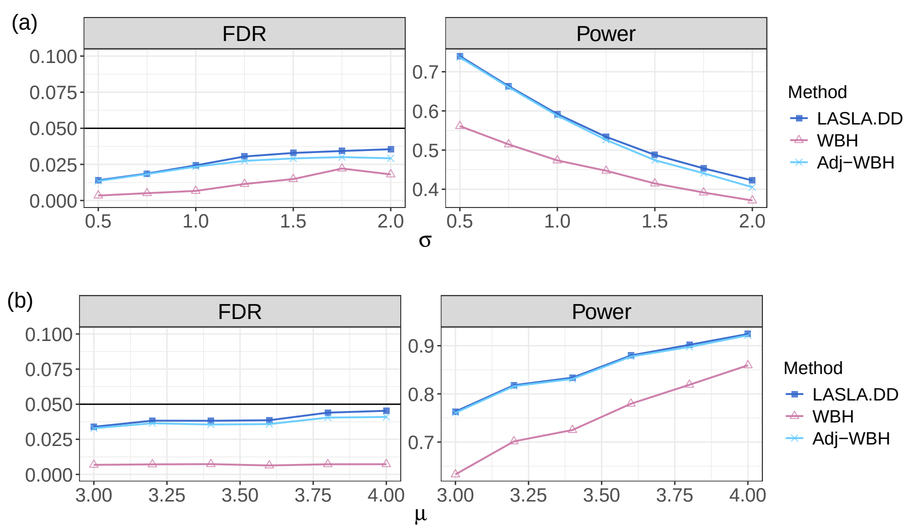

Figure S4.6: Empirical FDR and power comparison of different thresholding rules at nominal FDR level $\alpha = 0.05$ under the latent variable setting in Section S4.4. Other details are the same as in Figure S4.4.

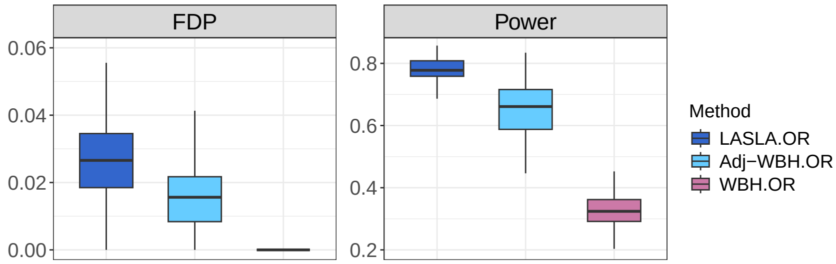

Figure S4.7: Empirical FDR and power comparison for the oracle LASLA, WBH and adjusted WBH for the heterogeneous sparsity level setting. All methods utilize the oracle LASLA weights.

S5 Numerical experiments for dependent data

In this section we conduct more numerical studies under data dependency. Following a similar setup as in Section 4, in all the subsequent experiments, $\theta_i \sim \text{Bernoulli}(0.1)$ indicates the presence or absence of a signal at index i. The primary statistics T_i are marginally distributed as $N(0,1)$ when $\theta_i = 0$ and are distributed as $N(3,1)$ when $\theta_i = 1$. The distance matrix $\boldsymbol{D} = (D_{ij})_{1\le i,j\le m}$ is defined by $D_{ij} \sim I_{\{\theta_i=\theta_j\}}|N(0,0.7)| + I_{\{\theta_i\neq\theta_j\}}|N(1,0.7)|$. We fix $m = 1000$ and the FDR level at $\alpha = 0.05$. The correlation structure will be specified in each setting below. Section S5.1 examines LASLA's performance in a weakly dependent setting, while Section S5.2 considers stronger dependency scenarios.

S5.1 Block dependency

In this setting, we consider a "block" dependency type where variables within the same block are equally correlated with each other, while variables in different blocks are uncorrelated. This structure offers a simplified model to mimic the scenarios where there are distinct clusters or groups of highly correlated variables that have little or no correlation with variables in other groups, similar

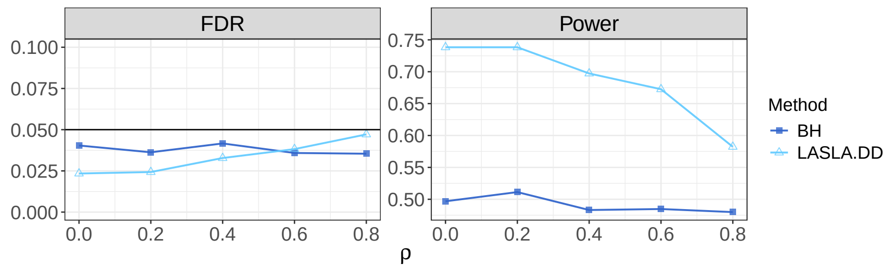

Figure S5.8: Empirical FDR and power comparison for data-driven LASLA and BH under the block dependency setting. The correlation strength increases as ρ increases.

to GWAS data where certain clusters of SNPs may work together to influence specific phenotypes.

We divide the $m = 1000$ indices into 10 blocks, each has a size of 100. For $k \in [10]$, let $b_k \subset [m]$ denote the collection of indices in block k. For $i, j \in [m]$, define the correlation matrix $\mathbf{\Sigma}$ as

$$\mathbf{\Sigma}_{ij} = \begin{cases} 1, & \text{if } i = j; \\ \rho, & \text{if } i \neq j, \text{ and } i, j \in b_k, \text{ for some } k \in [10]; \\ 0, & \text{otherwise}, \end{cases} \tag{S5.6}$$

where ρ controls the correlation strength. We vary ρ from 0 to 0.8 by step of 0.2, and summarize the result in Figure S5.8. We observe that LASLA's performance remains robust under block dependency, consistently controlling the FDR within the nominal level.

S5.2 Random dependency

In this section we consider a "random" dependency structure where the correlation matrix is generated randomly, with no specific pattern or clustering. We first generate a random factor vector $\boldsymbol{v} \in \mathbb{R}^m$, where each entry follows a standard normal distribution. Define the correlation matrix as

$$\mathbf{\Sigma}_{ij} = s(\boldsymbol{v}\boldsymbol{v}^\top + \boldsymbol{\epsilon}), \tag{S5.7}$$

where we add a diagonal matrix $\boldsymbol{\epsilon}$, with the diagonal elements uniformly distributed from $[0, 1]$, ensuring the positive definiteness of the correlation matrix. The function $s(\cdot)$ is then applied to standardize the matrix, ensuring that all diagonal elements are equal to 1. The off-diagonal elements approximately follow a uniform distribution from $[-1, 1]$. Unlike the block correlation in (S5.6) where indices are weakly and positively correlated, the random correlation in (S5.7) allows for both negative and positive correlations. Moreover, the correlation is strong and may even violate the weak dependency assumption in (A3). To adjust the dependency strength, we introduce a parameter a that scales the off-diagonal elements of $\boldsymbol{\Sigma}$ by dividing them by a. That is, for $i, j \in [m]$, the adjusted correlation matrix is defined as

$$\tilde{\boldsymbol{\Sigma}}_{ij} = \begin{cases} 1, & \text{if } i = j; \\ \frac{\boldsymbol{\Sigma}_{ij}}{a}, & \text{otherwise.} \end{cases}$$

We examine LASLA's performance with a taking values in $(1, 1.5, 2, 3, 5, 10)$. Note that smaller values of a signify stronger correlations. Figure S5.9 shows that the FDR of LASLA tends to rise above the nominal level under the strongest dependency setting. As discussed in Remark 2, the dependency assumption can be further relaxed by choosing a larger bandwidth, for instance $m^{-1/6}$. Hence, we rerun the experiment with $a = 1$ using the enlarged bandwidth, and Figure S5.10 demonstrates that LASLA effectively reduces the FDR level under this adjustment.

S6 Proof of Main Results

Recall that $\boldsymbol{D}_i$ is the ith column of $\boldsymbol{D}$, and $\mathcal{D}_i$ is a continuous finite domain (w.r.t. coordinate i) in $\mathbb{R}$ with positive measure by adopting the fixed-domain asymptotics in Stein (1995). Each $d \in \mathcal{D}_i$ is

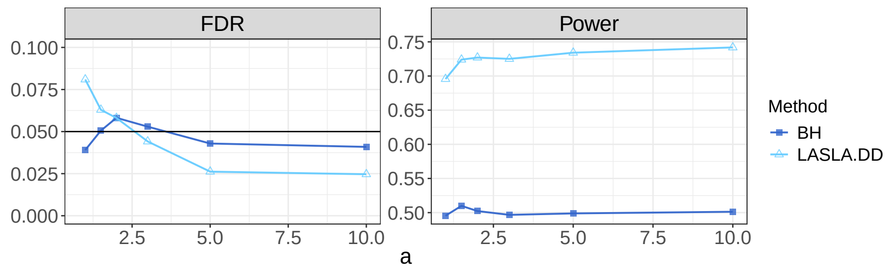

Figure S5.9: Empirical FDR and power comparison for data-driven LASLA and BH under the random dependency setting. The correlation strength decreases as a increases.

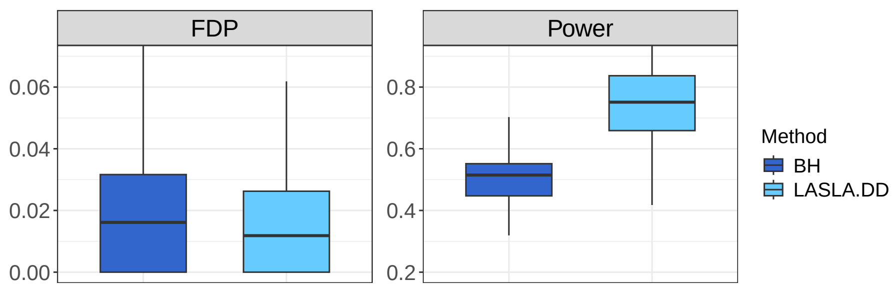

Figure S5.10: Empirical FDR and power comparison for data-driven LASLA and BH under the random dependency setting with $a = 1$ and enlarged bandwidth $h = m^{-1/6}$.

a distance and $0 \in \mathcal{D}_i$. The two sets $\boldsymbol{D}_i$ and $\mathcal{D}_i$ can be viewed as collections of distances measured from the partial and full network respectively, and it follows that $\boldsymbol{D}_i \subset \mathcal{D}_i$.

Throughout the proofs, we assume that $\boldsymbol{D}_i \to \mathcal{D}_i$ as $m \to \infty$ in the sense that, for any $d_0 \in \mathcal{D}_i$, there exists at least an index j such that $|D_{ij} - d_0| = O(m^{-1})$ as $m \to \infty$.

S6.1 Proof of Proposition 1

Proof. For simplicity of notation, throughout we omit the conditioning on $\mathcal{D}$, and use $\mathbb{P}(P_j > \tau | D_{ij} = x)$ and $\mathbb{P}(P_i > \tau)$ to denote $\mathbb{P}(P_j > \tau \mid \mathcal{D}_j, D_{ij} = x)$ and $\mathbb{P}(P_i > \tau \mid \mathcal{D}_i)$ respectively. Recall that

$$1-\pi_i = \frac{\sum_{j\in\mathcal{N}_i}[K_h(D_{ij})\mathbb{I}\{P_j>\tau\}]}{(1-\tau)\sum_{j\in\mathcal{N}_i}K_h(D_{ij})}$$

Also note that, for all $i\in[m]$,

$$\mathbb{E}\bigg(\sum_{j\in\mathcal{N}_i}[K_h(D_{ij})\mathbb{I}\{P_j>\tau\}\mid \boldsymbol{D}\bigg) = \sum_{j\in\mathcal{N}_i}[K_h(D_{ij})\,\mathbb{P}(P_j>\tau\mid \boldsymbol{D}_j)]$$

Then by $m^{-1}\ll h\ll m^{-\epsilon}$, as $\boldsymbol{D}_i\to\mathcal{D}_i$, we have

$$\frac{\mathbb{E}(1-\pi_i\mid \boldsymbol{D})}{\int_{\tilde{\mathcal{D}}_i}K_h(x)\,\mathbb{P}(P_{j_x}>\tau\mid D_{ij_x}=x)\,dx/(1-\tau)\int_{\tilde{\mathcal{D}}_i}K_h(x)\,dx}\to 1,$$

where j_x represents the index such that $D_{ij_x}=x$ and $\tilde{\mathcal{D}}_i$ is the limit of $\{D_{ij}, j\in\mathcal{N}_i\}$ in the asymptotic framework described at the beginning of Section S6. Using Taylor expansion at $x=0$, combined with Assumption (A2), we have

$$\begin{aligned}
&\int_{\tilde{\mathcal{D}}_i}K_h(x)\,\mathbb{P}(P_{j_x}>\tau\mid D_{ij_x}=x)\,dx\\
&=\mathbb{P}(P_i>\tau)\int_{\tilde{\mathcal{D}}_i}K_h(x)\,dx+\mathbb{P}'(P_i>\tau)\int_{\tilde{\mathcal{D}}_i}xK_h(x)\,dx\\
&+\frac{\mathbb{P}''(P_i>\tau)}{2}\int_{\tilde{\mathcal{D}}_i}x^2K_h(x)\,dx+O(h^2).
\end{aligned}$$

Thus, by the assumptions of $K(\cdot)$ in (2.4), uniformly for all index i, there exists some constant $c>0$ such that

$$\begin{aligned}
&[\mathbb{E}(\pi_i\mid \boldsymbol{D})-\pi_i^\tau]^2\\
&\le\bigg(c\int_{\tilde{\mathcal{D}}_i}xK_h(x)\,dx\bigg/\int_{\tilde{\mathcal{D}}_i}K_h(x)\,dx+c\int_{\tilde{\mathcal{D}}_i}x^2K_h(x)\bigg/\int_{\tilde{\mathcal{D}}_i}K_h(x)\,dx\bigg)^2+o(1)\\
&\to 0,\qquad\text{as } h\to 0.
\end{aligned}$$

Now we inspect the variance term. By Condition (A3), there exists a constant $c' > 1$,

$$\mathrm{Var}\Big(\sum_{j\in\mathcal{N}_i}[K_h(D_{ij})\mathbb{I}\{P_j > \tau\}] \mid \boldsymbol{D}\Big)$$

$$\le c' \sum_{j\in\mathcal{N}_i}[K_h^2(D_{ij})\,\mathrm{Var}\big(\mathbb{I}\{P_j > \tau\} \mid \boldsymbol{D}_j\big)]$$

$$= c' \sum_{j\in\mathcal{N}_i}\big[K_h^2(D_{ij})\,\mathbb{P}(P_j > \tau \mid \boldsymbol{D}_j)\{1 - \mathbb{P}(P_j > \tau \mid \boldsymbol{D}_j)\}\big].$$

Hence, as $h \gg m^{-1}$, by the assumptions of $K(\cdot)$ in (2.4) and that it is positive and bounded, we have

$$\begin{aligned}
\mathrm{Var}(1-\pi_i) &\le c'' m^{-1}\frac{\int_{\tilde{\mathcal{D}}_i} K_h^2(x)\,dx}{[(1-\tau)\int_{\tilde{\mathcal{D}}_i} K_h(x)\,dx]^2}\\
&\le c''(mh)^{-1}\frac{\int_{\mathbb{R}} K^2(y)\,dy}{[(1-\tau)\int_{\tilde{\mathcal{D}}_i} K_h(x)\,dx]^2}\\
&\le c'''(mh)^{-1} = o(1),
\end{aligned}$$

for some constant $c'', c''' > 0$. Hence, as $\boldsymbol{D}_i \to \mathcal{D}_i$, by combining the bias term and variance term, the consistency result is proved. □

S6.2 Proof of Theorem 1

Proof. For simplicity of notation, throughout we omit the explicit conditioning on $\mathcal{D}$, and use $\mathbb{P}(\theta_i = 0)$ to denote $\mathbb{P}(\theta_i = 0 \mid \mathcal{D}_i)$ and $\mathbb{P}(P_i^w \le t \mid \theta_i = 0, w_i)$ to denote $\mathbb{P}(P_i^w \le t \mid \theta_i = 0, w_i, \mathcal{D}_i)$. Note that, by Algorithm 1, the FDP of LASLA at the thresholding level t can be calculated by

$$\begin{aligned}
FDP(t) &= \frac{\sum_{i=1}^m \mathbb{I}\{P_i^w \le t, \theta_i = 0\}}{\max[\sum_{i=1}^m \mathbb{I}\{P_i^w \le t\}, 1]}\\
&= \frac{\sum_{i=1}^m \mathbb{P}(P_i^w \le t \mid \theta_i = 0, w_i)\mathbb{P}(\theta_i = 0)}{\max[\sum_{i=1}^m \mathbb{I}\{P_i^w \le t\}, 1]} \cdot \frac{\sum_{i=1}^m \mathbb{I}\{P_i^w \le t, \theta_i = 0\}}{\sum_{i=1}^m \mathbb{P}(P_i^w \le t \mid \theta_i = 0, w_i)\mathbb{P}(\theta_i = 0)},
\end{aligned}$$

Step 1: We first show that, uniformly for all $i \in [m]$, we have

$$\sum_{i=1}^{m} \mathbb{P}(P_i^w \le t \mid \theta_i = 0, w_i)\mathbb{P}(\theta_i = 0) \le [1 + o_{\mathbb{P}}(1)] \sum_{i=1}^{m} w_i(1 - \pi_i)t. \tag{S6.8}$$

Note that, in Algorithm 2, T_i is not used in the computation of w_i given the sign of T_i. Then by the independence assumption (A1), T_i is independent of w_i conditioning on the sign, and it follows that:

$$\begin{aligned}
&\sum_{i=1}^{m} \mathbb{P}(P_i^w \le t \mid \theta_i = 0, w_i)\mathbb{P}(\theta_i = 0)\\
&= \left[\sum_{i=1}^{m} \mathbb{P}(P_i^w \le t \mid T_i > 0, \theta_i = 0, w_i)\mathbb{P}(T_i > 0 \mid \theta_i = 0, w_i)(1 - \pi_i^*)\right]\\
&\quad + \left[\sum_{i=1}^{m} \mathbb{P}(P_i^w \le t \mid T_i < 0, \theta_i = 0, w_i)\mathbb{P}(T_i < 0 \mid \theta_i = 0, w_i)(1 - \pi_i^*)\right]\\
&= \left[\sum_{i=1}^{m} \mathbb{P}(T_i > 0 \mid \theta_i = 0)w_i(1 - \pi_i^*)t\right] + \left[\sum_{i=1}^{m} \mathbb{P}(T_i < 0 \mid \theta_i = 0)w_i(1 - \pi_i^*)t\right]\\
&= \sum_{i=1}^{m} w_i(1 - \pi_i^*)t \le \sum_{i=1}^{m} w_i(1 - \pi_i^\tau)t,
\end{aligned}$$

where the last inequality follows from the fact that π_i^τ is a conservative approximation of π_i^* as showed in Cai et al. (2022). By the result of Proposition 1 and Assumption (A4), together with the fact that $\xi \le w_i \le 1$ for $i \in [m]$, we have

$$\sum_{i=1}^{m} \mathbb{P}(P_i^w \le t \mid \theta_i = 0, w_i)\mathbb{P}(\theta_i = 0) \le \sum_{i=1}^{m} w_i[1 - \pi_i + o_{\mathbb{P}}(1)]t = [1 + o_{\mathbb{P}}(1)] \sum_{i=1}^{m} w_i(1 - \pi_i)t.$$

Hence, (S6.8) is proved.

Step 2: We next show that

$$\left| \frac{\sum_{\theta_i=0} \mathbb{P}(P_i^w \le t \mid \theta_i = 0, w_i) - \sum_{i=1}^{m} \mathbb{P}(P_i^w \le t \mid \theta_i = 0, w_i)\mathbb{P}(\theta_i = 0)}{\sum_{i=1}^{m} \mathbb{P}(P_i^w \le t \mid \theta_i = 0, w_i)\mathbb{P}(\theta_i = 0)} \right| \to 0, \tag{S6.9}$$

in probability. Define the event

$$B = \left[\{\theta_i\}_{i=1}^m, \sum_{i=1}^m \mathbb{I}\{\theta_i = 0\} \geq cm \text{ for some constant } c > 0\right].$$

It follows from Condition (A4) that $\mathbb{P}(B) \to 1$. Then by the fact that $\xi \leq w_i \leq 1$, we have

$$\begin{aligned}
&\mathbb{E}\left[\mathbb{E}\left(\left|\frac{\sum_{i=1}^m [\mathbb{P}(P_i^w \leq t \mid \theta_i = 0, w_i)\mathbb{I}\{\theta_i = 0\} - \mathbb{P}(P_i^w \leq t \mid \theta_i = 0, w_i)\mathbb{P}(\theta_i = 0)]}{\sum_{i=1}^m \mathbb{P}(P_i^w \leq t \mid \theta_i = 0, w_i)\mathbb{P}(\theta_i = 0)}\right|^2 \middle| w_i\right)\right] \\
&= \mathbb{E}\left[\mathbb{E}\left(\left|\frac{\sum_{i=1}^m \mathbb{P}(P_i^w \leq t \mid \theta_i = 0, w_i)[\mathbb{I}\{\theta_i = 0\} - \mathbb{P}(\theta_i = 0)]}{\sum_{i=1}^m \mathbb{P}(P_i^w \leq t \mid \theta_i = 0, w_i)\mathbb{P}(\theta_i = 0)}\right|^2 \middle| w_i\right)\right] \\
&= \mathbb{E}\left[\mathrm{Var}\left(\sum_{i=1}^m [\mathbb{P}(P_i^w \leq t | \theta_i = 0)\mathbb{I}\{\theta_i = 0\}] \middle| w_i\right) \Big/ \left(\sum_{i=1}^m w_i t \mathbb{P}(\theta_i = 0)\right)^2\right] \\
&= O(m^{\zeta - 1}),
\end{aligned}$$

the last equality follows from the law of total variance and condition (A3) for some $0 \leq \zeta < 1$. Hence (S6.9) is proved.

Step 3: Finally, we analyze the following quantity:

$$\frac{\sum_{i=1}^m \mathbb{I}\{P_i^w \leq t, \theta_i = 0\}}{\sum_{i=1}^m \mathbb{P}(P_i^w \leq t \mid \theta_i = 0, w_i)\mathbb{P}(\theta_i = 0)}.$$

We first check the range of the cutoff t, or equivalently the threshold for the weighted z-values, i.e., $z_i^w = \Phi^{-1}(1 - P_i^w/2)$, for $i \in [m]$. Then as shown in Cai et al. (2022) and replace their weights $\frac{\pi_i}{1-\pi_i}$ by w_i, it is easy to see that, by applying BH procedure at level α to the adjusted p-values with weights w_i, the corresponding threshold is no larger than the threshold of LASLA for the adjusted p-values with the same weights w_i. Hence it suffices to obtain the threshold for the weighted z-values $z_i^w = \Phi^{-1}(1 - P_i^w/2)$ of such BH procedure with weights w_i.

Let $t_m = (2\log m - 2\log\log m)^{1/2}$. By Condition (A5), we have

$$\sum_{\theta_i=1} \mathbb{I}\{|z_i| \geq (c\log m)^{1/2+\rho/4}\} \geq \{1/(\pi^{1/2}\alpha)+\delta\}(\log m)^{1/2},$$

with probability going to one. Recall that we have $\xi \leq w_i \leq 1$ for some constant $\xi > 0$. Thus, for those indices $i \in \mathcal{H}_1$ (equivalently $\theta_i = 1$) such that $|z_i| \geq (c\log m)^{1/2+\rho/4}$, we have

$$P_i^w \leq (1-\Phi((c\log m)^{1/2+\rho/4}))/w_i = o(m^{-M}),$$

for any constant $M > 0$. Thus we have

$$\sum_{i\in[m]} \mathbb{I}\{z_i^w \geq (2\log m)^{1/2}\} \geq \{1/(\pi^{1/2}\alpha)+\delta\}(\log m)^{1/2},$$

with probability going to one. Hence, with probability tending to one,

$$\frac{2m}{\sum_{i\in[m]} \mathbb{I}\{z_i^w \geq (2\log m)^{1/2}\}} \leq 2m\{1/(\pi^{1/2}\alpha)+\delta\}^{-1}(\log m)^{-1/2}.$$

Because $1-\Phi(t_m) \sim 1/\{(2\pi)^{1/2}t_m\}\exp(-t_m^2/2)$, it suffices to show that,

$$\sup_{0\leq t\leq t_m}\left|\frac{\sum_{i=1}^m \mathbb{I}\{z_i^w \geq t, \theta_i = 0\} - \sum_{i=1}^m \mathbb{P}(z_i^w \geq t \mid \theta_i = 0, w_i)\mathbb{P}(\theta_i = 0)}{\sum_{i=1}^m \mathbb{P}(z_i^w \geq t \mid \theta_i = 0, w_i)\mathbb{P}(\theta_i = 0)}\right| \to 0, \tag{S6.10}$$

in probability. Let the event $A = [\{\theta_i\}_{i=1}^m : \text{(S6.9) holds}]$. By the proofs in Step 2, we have $\mathbb{P}(A) \to 1$. Hence, it is enough to show that, for $\{\theta_i\}_{i=1}^m \in A$, we have

$$\sup_{0\leq t\leq t_m}\left|\frac{\sum_{\theta_i=0}[\mathbb{I}\{z_i^w \geq t\} - \mathbb{P}(z_i^w \geq t \mid \theta_i = 0, w_i)]}{\sum_{\theta_i=0}\mathbb{P}(z_i^w \geq t \mid \theta_i = 0, w_i)}\right| \to 0, \tag{S6.11}$$

in probability. Let $0 \leq t_0 < t_1 < \cdots < t_b = t_m$ such that $t_\iota - t_{\iota-1} = v_m$ for $1 \leq \iota \leq b-1$ and $t_b - t_{b-1} \leq v_m$, where $v_m = 1/\sqrt{\log m(\log_4 m)}$. Thus we have $b \sim t_m/v_m$. For any t such that $t_{\iota-1} \leq t \leq t_\iota$, due to the fact that $G(t + o((\log m)^{-1/2}))/G(t) = 1 + o(1)$ with $G(t) = 2(1-\Phi(t))$

uniformly in $0 \le t \le c(\log m)^{1/2}$ for any constant c, by Xia et al. (2020), it suffices to prove that

$$\max_{0\le\iota\le b}\left|\frac{\sum_{\theta_i=0}[\mathbb{I}\{z_i^w \ge t_\iota\} - \mathbb{P}(z_i^w \ge t_\iota \mid \theta_i = 0, w_i)]}{\sum_{\theta_i=0}\mathbb{P}(z_i^w \ge t_\iota \mid \theta_i = 0, w_i)}\right| \to 0, \tag{S6.12}$$

in probability. Thus, it suffices to show that, for any $\epsilon > 0$,

$$\int_0^{t_m}\mathbb{P}\left\{\left|\frac{\sum_{\theta_i=0}[\mathbb{I}\{z_i^w \ge t\} - \mathbb{P}(z_i^w \ge t \mid \theta_i = 0, w_i)]}{\sum_{\theta_i=0}\mathbb{P}(z_i^w \ge t \mid \theta_i = 0, w_i)}\right| \ge \epsilon\right\}dt = o(v_m). \tag{S6.13}$$

By the fact that $\xi \le w_i \le 1$ for some constant $\xi > 0$, we have

$$\begin{aligned}
\mathbb{P}(z_i^w \ge t \mid \theta_i = 0, w_i) &= \mathbb{P}(\Phi^{-1}(1 - P_i^w/2) \ge t \mid \theta_i = 0, w_i) \\
&= \mathbb{P}(P_i \le 2w_i(1 - \Phi(t)) \mid \theta_i = 0, w_i) \\
&= 2w_i(1 - \Phi(t)) \ge \xi G(t).
\end{aligned}$$

It follows that

$$\begin{aligned}
&\mathbb{E}\left|\frac{\sum_{\theta_i=0}[\mathbb{I}\{z_i^w \ge t\} - \mathbb{P}(z_i^w \ge t \mid \theta_i = 0, w_i)]}{\sum_{\theta_i=0}\mathbb{P}(z_i^w \ge t \mid \theta_i = 0, w_i)}\right|^2 \\
&\le \frac{\mathbb{E}\left|\sum_{\theta_i=0}[\mathbb{I}\{z_i^w \ge t\} - \mathbb{P}(z_i^w \ge t \mid \theta_i = 0, w_i)]\right|^2}{\left\{\sum_{\theta_i=0}\xi G(t)\right\}^2} \\
&= \frac{\mathbb{E}\Big[\sum_{\theta_i=0,\theta_j=0}\mathbb{P}(z_i^w \ge t, z_j^w \ge t \mid \theta_i = 0, \theta_j = 0, w_i, w_j) - \left\{\sum_{\theta_i=0}\mathbb{P}(z_i^w \ge t \mid \theta_i = 0, w_i)\right\}^2\Big]}{\left\{\sum_{\theta_i=0}\xi G(t)\right\}^2}.
\end{aligned}$$

Recall that, by Algorithm 2 we only use $O(m^{1-\epsilon})$ neighbors to construct w_i for any small enough constant $\epsilon > 0$, Hence, we can divide the indices pairs $\tilde{\mathcal{H}}_0 = \{(i,j) : \theta_i = 0, \theta_j = 0\}$ into two subsets:

$$\begin{aligned}
\tilde{\mathcal{H}}_{01} &= \{(i,j) \in \tilde{\mathcal{H}}_0, \text{ either } P_i^w \text{ is correlated with } P_j \text{ or } P_j^w \text{ is correlated with } P_i\}, \\
\tilde{\mathcal{H}}_{02} &= \tilde{\mathcal{H}}_0 \setminus \tilde{\mathcal{H}}_{01},
\end{aligned}$$

where $|\tilde{\mathcal{H}}_{01}| = O(m^{2-\epsilon})$ while among them m pairs with $i = j$ are perfectly correlated.

Note that, for $(i,j) \in \tilde{\mathcal{H}}_{01}$,

$$\frac{\mathbb{E}\Big[\sum_{(i,j)\in\tilde{\mathcal{H}}_{01}}\Big\{\mathbb{P}(z_i^w \geq t, z_j^w \geq t \mid \theta_i = 0, \theta_j = 0, w_i, w_j) - \prod_{h=i,j}\mathbb{P}(z_h^w \geq t \mid \theta_h = 0, w_h)\Big\}\Big]}{\big\{\sum_{\theta_i=0}\xi G(t)\big\}^2}$$
$$\leq \frac{\mathbb{E}\Big[\sum_{(i,j)\in\tilde{\mathcal{H}}_{01}}\mathbb{P}(z_i^w \geq t, z_j^w \geq t \mid \theta_i = 0, \theta_j = 0, w_i, w_j)\Big]}{\big\{\sum_{\theta_i=0}\xi G(t)\big\}^2}.$$

Recall that event $B = [\{\theta_i\}_{i=1}^m, \sum_{i=1}^m \mathbb{I}\{\theta_i = 0\} \geq cm$ for some constant $c > 0]$ and $\mathbb{P}(B) \to 1$. For $\{\theta_i\}_{i=1}^m \in A \cap B$, we have

$$\frac{\mathbb{E}\Big[\sum_{(i,j)\in\tilde{\mathcal{H}}_{01}}\Big\{\mathbb{P}(z_i^w \geq t, z_j^w \geq t \mid \theta_i = 0, \theta_j = 0, w_i, w_j) - \prod_{h=i,j}\mathbb{P}(z_h^w \geq t \mid \theta_h = 0, w_h)\Big\}\Big]}{\big\{\sum_{\theta_i=0}\xi G(t)\big\}^2}$$
$$\leq \frac{\sum_{\theta_i=0}\mathbb{P}(z_i^w \geq t \mid \theta_i = 0, w_i)}{\big\{\sum_{\theta_i=0}\xi G(t)\big\}^2} + O\left(\frac{m^{2-\epsilon}}{m^2}\right)$$
$$\leq O\left(\frac{1}{mG(t)}\right) + O\left(m^{-\epsilon}\right),$$

where the first term reflects the pairs with $i = j$. On the other hand,

$$\mathbb{E}\left[\sum_{(i,j)\in\tilde{\mathcal{H}}_{02}}\left\{\mathbb{P}\left\{z_i^w \geq t, z_j^w \geq t \mid \theta_i = 0, \theta_j = 0, w_i, w_j\right\} - \prod_{h=i,j}\mathbb{P}(z_h^w \geq t \mid \theta_h = 0, w_h)\right\}\right] = 0.$$

Then by the fact that

$$\int_0^{t_m}\left\{\frac{1}{mG(t)} + m^{-\epsilon}\right\}dt = o(v_m),$$

and that $\mathbb{P}(A \cap B) \to 1$, (S6.13) is proved and (S6.11) is thus proved. Combining (S6.11) and (S6.9), we obtain (S6.10). This together with (S6.8) prove the result of Theorem 1. □

S6.3 Proof of Theorem 2

Proof. Note that

$$Q^1(t) = \frac{\sum_{i=1}^m (1 - \pi_i^*)t}{\sum_{i=1}^m (1 - \pi_i^*)t + \sum_{i=1}^m \pi_i^* F_{1i}^*(t \mid \mathcal{D})}.$$

Recall that $\tilde{w}_i = w_i[\sum_{j=1}^m (1-\pi_j^*)]/[\sum_{j=1}^m (1-\pi_j^*)w_j]$, we have

$$Q^{\tilde{w}}(t) = \frac{\sum_{i=1}^m (1-\pi_i^*)\tilde{w}_i t}{\sum_{i=1}^m (1-\pi_i^*)\tilde{w}_i t + \sum_{i=1}^m \pi_i^* F_{1i}^*(\tilde{w}_i t \mid \mathcal{D})}$$
$$= \frac{\sum_{i=1}^m (1-\pi_i^*)t}{\sum_{i=1}^m (1-\pi_i^*)t + \sum_{i=1}^m \pi_i^* F_{1i}^*(\tilde{w}_i t \mid \mathcal{D})}.$$

Under the Assumption (A7) we have,

$$\sum_{i=1}^m \pi_i^* F_{1i}^*(\tilde{w}_i t \mid \mathcal{D}) = \sum_{i=1}^m \pi_i^* F_{1i}^*(t/\tilde{w}_i^{-1} \mid \mathcal{D})$$
$$\geq \sum_{i=1}^m \pi_i^* F_{1i}^* \left(\frac{\sum_{i=1}^m \pi_j^* t}{\sum_{j=1}^m \pi_j^* \tilde{w}_j^{-1}} \mid \mathcal{D} \right).$$

By Assumption (A6) and the construction that $\tilde{w}_i = w_i[\sum_{j=1}^m (1-\pi_j^*)]/[\sum_{j=1}^m (1-\pi_j^*)w_j]$, we have $[\sum_{i=1}^m \pi_i^*]/\left[\sum_{i=1}^m \pi_i^* \tilde{w}_i^{-1}\right] \geq 1$. Therefore,

$$\sum_{i=1}^m \pi_i^* F_{1i}^*(\tilde{w}_i t \mid \mathcal{D}) \geq \sum_{i=1}^m \pi_i^* F_{1i}^*(t \mid \mathcal{D}).$$

Hence, by the definition of t_o^1, it is easy to see that $Q^{\tilde{w}}(t_o^1) \leq Q^1(t_o^1) \leq \alpha$. It yields that $t_o^{\tilde{w}} \geq t_o^1$ and thus $\Psi^{\tilde{w}}(t_o^{\tilde{w}}) \geq \Psi^{\tilde{w}}(t_o^1) \geq \Psi^1(t_o^1)$. $\square$

S7 Asymptotic theories under weak dependence

In this section, we study the asymptotic control of FDP and FDR for dependent p-values. We collect some additional regularity conditions to develop the theories under weak dependence. We first introduce in Section S7.1 the benchmark oracle weight. Then the proofs are developed in two stages: Section S7.2 shows the consistency of the weight estimators; Section S7.3 illustrates that the oracle-assisted LASLA controls FDP and FDR asymptotically.

S7.1 Oracle weight

With slight abuse of notation, we let $L_i^* = (1-\pi_i^\tau)f_0(t_i)/f_i^*(t_i|\mathcal{D})$, where $f_i^*(\cdot|\mathcal{D})$ can be interpreted as the density function of the primary statistic in light of the full network. Again we omit the conditioning on $\mathcal{D}$ throughout for notation simplicity. Since $f_i(t)$ is calculated in light of the partial network $\boldsymbol{D}_i$, it should become close to $f_i^*(t)$ as $\boldsymbol{D}_i \to \mathcal{D}_i$, which will be shown rigorously later in Section S7.2.

Similarly as the oracle-assisted weights defined in Section 2.3, denote the sorted statistics by $L_{(1)}^* \leq \ldots \leq L_{(m)}^*$. Let $L_{(k^*)}^*$ be the threshold, where $k^* = \max\{j : j^{-1}\sum_{i=1}^{j} L_{(i)}^* \leq \alpha\}$. Then for $T_i > 0$, let $t_i^{*,+} = \infty$ if $(1-\pi_i^*)f_0(t)/f_i^*(t)\} \geq \mathbb{E}\left\{L_{(k^*)}^*\right\}$ for all $t \geq 0$, else:

$$t_i^{*,+} = \inf\left[t \geq 0 : \{(1-\pi_i^\tau)f_0(t)/f_i^*(t)\} \leq \mathbb{E}\left\{L_{(k^*)}^*\right\}\right],$$

and define $w_i^* = 1 - F_0(t_i^{*,+})$. For $T_i < 0$, we let $t_i^{*,-} = -\infty$ if $(1-\pi_i^*)f_0(t)/f_i^*(t)\} \geq \mathbb{E}\left\{L_{(k^*)}^*\right\}$ for all $t \leq 0$, else:

$$t_i^{*,-} = \sup\left[t \leq 0 : \{(1-\pi_i^\tau)f_0(t)/f_i^*(t)\} \leq \mathbb{E}\left\{L_{(k^*)}^*\right\}\right],$$

and the corresponding weight is given by $w_i^* = F_0(t_i^{*,-})$. Again, we let $w_i^* = \max\{w_i^*, \xi\}$ and $w_i^* = \min\{w_i^*, 1-\xi\}$ for any sufficiently small constant $0 < \xi < 1$. Then the oracle thresholding rule is provided by

$$k^{*,w} = \max\left\{j : (1/j)\sum_{i=1}^{m} w_i^*(1-\pi_i^\tau)P_{(j)}^{w^*} \leq \alpha\right\}. \tag{S7.14}$$

We show next that the oracle-assisted weight w_i in Algorithm 1 estimates w_i^* consistently under some regularity conditions in the following section.

S7.2 Consistency of the weight estimator

The weight consistency result is built upon the consistency of sparsity estimator (2.5) and density estimator (2.6). The theoretical properties of the former can be similarly proved as Proposition 1 under conditions (A2) and (A3), while letting $\mathcal{N}_i = \{j \in [m], j \neq i\}$ and $h \gg m^{-1}$. We shall focus on the consistency of the density estimator below. Recall that

$$f_i(t) = \frac{\sum_{j\neq i}[V_h(i,j)K_h(t_j - t)]}{\sum_{j\neq i} V_h(i,j)}.$$

We will focus on the cases when the support of the primary statistics $\boldsymbol{T} = \{T_i : i \in [m]\}$ is $\mathbb{R}$, e.g. z-statistics and t-statistics.

(A8) Assume that for all i, j, $f_j^*(t \mid D_{ij} = x)$ has bounded first and second partial derivatives at t and x.

(A9) Assume that, for all $i \in [m]$,

$$\text{Var}\left\{\sum_{j=1}^{m} K_h(D_{ij})K_h(t_j - t) \mid \boldsymbol{D}\right\} \leq C \sum_{j=1}^{m} \text{Var}\left\{K_h(D_{ij})K_h(t_j - t) \mid \boldsymbol{D}_j\right\}$$

for some constant $C > 1$, for all t.

Remark 1. Assumption (A8) is a mild regularity condition on the densities of the primary statistics. Condition (A9) assumes that most of the primary statistics are weakly correlated.

Lemma 1. *Let $K(\cdot)$ be a kernel function that satisfies (2.4) and let T be a random variable with support $\mathbb{R}$. Assume that its conditional density $f(\cdot \mid \boldsymbol{D})$ has bounded first and second derivatives.*

Then for any fixed t, as the bandwidth $h \to 0$, we have

$$\mathbb{E}(K_h(T-t) \mid \boldsymbol{D}) = f(t \mid \boldsymbol{D}) + O(h^2)\sigma_K^2$$

$$\mathbb{E}(K_h^2(T-t) \mid \boldsymbol{D}) = f(t \mid \boldsymbol{D})\frac{R(K)}{h} + O(h)G(K),$$

where $R(K) = \int_{\mathbb{R}} K^2(x)\,dx$ and $G(K) = \int_{\mathbb{R}} x^2 K^2(x)\,dx$.

Once Lemma 1 is developed, we can obtain the following proposition on density estimation consistency.

Proposition 1. *Under Assumptions (A8) and (A9), if $h \gg m^{-1/2}$, we have for any t, uniformly for all $i \in [m]$,*

$$\mathbb{E}(\{f_i(t) - f_i^*(t)\} \mid \boldsymbol{D})^2 \to 0, \quad \textit{as } \boldsymbol{D}_i \to \mathcal{D}_i.$$

Next, we develop the consistency result of the oracle-assisted weight in Algorithm 1. Without loss of generality, we assume that $-\infty < t_i^{*,-} \le t_i^{*,+} < +\infty$ for all $i \in [m]$. Let $g_i(t) = (1 - \pi_i^\tau) f_0(t)/f_i^*(t)$ and define functions $g_{i,+}^{-1} : x \to t$ and $g_{i,-}^{-1} : x \to t$ as

$$g_{i,+}^{-1}(x) = \inf\{t \ge 0 : g_i(t) \le x\},$$

and

$$g_{i,-}^{-1}(x) = \sup\{t \le 0 : g_i(t) \le x\}.$$

We let $g_{i,+}^{-1}(x) = +\infty$ if $g_i(t) \ge x$ for all $t \ge 0$ and let $g_{i,-}^{-1}(x) = -\infty$ if $g_i(t) \ge x$ for all $t \le 0$. We also assume that $-\infty < t_i^- \le t_i^+ < +\infty$ for all $i \in [m]$ for simplicity. If not, the data-driven testing procedure will be more conservative than the oracle one and hence the asymptotic FDR control can again be guaranteed. Then based on Proposition 1, we obtain the following corollary.

Corollary 1. *Assume that* $g_{i,+}^{-1}(x)$ *and* $g_{i,-}^{-1}(x)$ *have bounded first derivative for all* $0 < x < 1$ *such that* $-\infty < g_{i,-}^{-1}(x) \le g_{i,+}^{-1}(x) < +\infty$ *and there exists some constants* α_1 *and* α_2 *such that* $\frac{1}{k^*}\sum_{i=1}^{k^*} L_{(i)}^* \le \alpha_1 < \alpha < \alpha_2 \le \frac{1}{k^*+1}\sum_{i=1}^{k^*+1} L_{(i)}^*$ *with probability tending to 1. Assume that* $1/f_i^*(t)$ *are bounded with probability tending to 1 uniformly for all* $i \in [m]$*. Further assume that* $\pi_i^\tau \le 1-\xi$ *for sufficiently small constant* $\xi > 0$ *and* $Var\left(L_{(k^*)}^*\right) = o(1)$*. Then under the conditions of Propositions 1 and 1, we have, as* $m \to \infty$, $w_i = w_i^* + o_{\mathbb{P}}(1)$, *uniformly for all* $i \in [m]$.

Remark 2. The conditions on $g_{i,+}^{-1}$, $g_{i,-}^{-1}$ and $L_{(i)}^*$ are mild and can be easily satisfied by the commonly used distributions such as normal distribution, t-distribution, etc. The condition on $1/f_i^*(t)$ can be further relaxed by a more sophisticated calculation on the convergence rate of $f_i(t)$ in the proof of Proposition 1. The condition $\mathrm{Var}\left(L_{(k^*)}^*\right) = o(1)$ is mild and can be satisfied by most of the settings in the scope of this paper. For example, in Setting 1 of Section S4.4, $\mathrm{Var}\left(L_{(k^*)}^*\right)$ is of the order 10^{-2}.

S7.3 FDP and FDR control under weak dependence

Recall that we define the z-values by $Z_i = \Phi^{-1}(1 - P_i/2)$, and let $\boldsymbol{Z} = (Z_1, \ldots, Z_m)^\intercal$. We collect below one additional regularity condition for the asymptotic error rates control. We allow dependency to come from two sources: Dependence of the θ_i's and dependency of the p-values given θ_i's. Our conditions on these two types of correlations are respectively specified in (A4) and (A10).

(A10) Define $(r_{i,j})_{m\times m} = \boldsymbol{R} = \mathsf{Corr}(\boldsymbol{Z})$. Assume $\max_{1\le i<j\le m} |r_{i,j}| \le r < 1$ for some constant $r > 0$. Moreover, there exists $\gamma > 0$ such that $\max_{\{i:\theta_i=0\}} |\Gamma_i(\gamma)| = o(m^\kappa)$ for some constant $0 < \kappa < \frac{1-r}{1+r}$, where $\Gamma_i(\gamma) = \{j : 1 \le j \le m, |r_{i,j}| \ge (\log m)^{-2-\gamma}\}$.

We first consider the oracle case. Recall that

$$k^{*,w} = \max\left\{ j : \left(P_{(j)}^{w^*}/j \right) \sum_{i=1}^{m} w_i^*(1-\pi_i^\tau) \le \alpha \right\}.$$

Denote the corresponding threshold for the weighted p-values as t^{w^*} and the set of decision rules as $\boldsymbol{\delta}^{w^*}$. The next theorem shows that both FDP and FDR are controlled at the nominal level asymptotically under dependency.

Theorem 1. *Under (A4), (A5) and (A10), we have for any* $\varepsilon > 0$,

$$\overline{\lim_{\boldsymbol{D}_i \to \mathcal{D}_i, \forall i}} FDR(\boldsymbol{\delta}^{w^*}) \le \alpha, \text{ and } \lim_{\boldsymbol{D}_i \to \mathcal{D}_i, \forall i} \mathbb{P}(FDP(\boldsymbol{\delta}^{w^*}) \le \alpha + \varepsilon) = 1.$$

The next theorem establishes the theoretical properties of the data-driven LASLA procedure. Recall that $\boldsymbol{\delta}^w \equiv \boldsymbol{\delta}^w(t^w) = \{\delta_i^w(t^w) : i \in [m]\}$ is the set of data-driven decision rules, where the LASLA weights are computed by Algorithm 1 with $\mathcal{N}_i = \{j \in [m], j \ne i\}$. Based on the weight consistency result, the FDP and FDR of data-driven LASLA can be asymptotically controlled under dependency.

Theorem 2. *Under the conditions in Corollary 1 and Theorem 1, we have for any* $\varepsilon > 0$,

$$\overline{\lim_{\boldsymbol{D}_i \to \mathcal{D}_i, \forall i}} FDR(\boldsymbol{\delta}^{w}) \le \alpha, \text{ and } \lim_{\boldsymbol{D}_i \to \mathcal{D}_i, \forall i} \mathbb{P}(FDP(\boldsymbol{\delta}^{w}) \le \alpha + \varepsilon) = 1.$$

S8 Proof of the theoretical results under dependency

S8.1 Proof of Lemma 1

Proof. By Taylor expansion of $f(y \mid \boldsymbol{D})$ at $y = t$, we have

$$\begin{aligned}
\mathbb{E}(K_h(T-t) \mid \boldsymbol{D}) &= \int K_h(y-t) f(y \mid \boldsymbol{D})\, dy \\
&= \int K_h(y-t)\left[f(t \mid \boldsymbol{D}) + f'(t \mid \boldsymbol{D})(y-t) + \frac{f''(t \mid \boldsymbol{D})}{2}(y-t)^2\right] dy + O(h^2) \\
&= f(t \mid \boldsymbol{D}) + O(h^2)\sigma_K^2.
\end{aligned}$$

Similarly,

$$\begin{aligned}
\mathbb{E}(K_h^2(T-t) \mid \boldsymbol{D}) &= \int K_h^2(y-t) f(y \mid \boldsymbol{D})\, dy \\
&= \int K_h^2(y-t)\left[f(t \mid \boldsymbol{D}) + f'(t \mid \boldsymbol{D})(y-t) + \frac{f''(t \mid \boldsymbol{D})}{2}(y-t)^2\right] dy + O(h) \\
&= \frac{f(t \mid \boldsymbol{D}) R(K)}{h} + O(h) G(K).
\end{aligned}$$

□

S8.2 Proof of Proposition 1

Proof. By Lemma 1, we have

$$\mathbb{E}(f_i(t) \mid \boldsymbol{D}) = \frac{\sum_{j \neq i} K_h(D_{ij})\, \mathbb{E}(K_h(t_j - t) \mid \boldsymbol{D}_j)}{\sum_{j \neq i} K_h(D_{ij})} = \frac{\sum_{j \neq i} K_h(D_{ij}) f_j^*(t \mid \boldsymbol{D}_j)}{\sum_{j \neq i} K_h(D_{ij})} + O(h^2).$$

By $h \gg m^{-1/2}$, as $\boldsymbol{D}_i \to \mathcal{D}_i$, we have

$$\frac{\sum_{j \neq i} K_h(D_{ij}) f_j^*(t \mid \boldsymbol{D}_j) / \sum_{j \neq i} K_h(D_{ij})}{\int_{\mathcal{D}_i} K_h(x) f_{j_x}^*(t \mid D_{i j_x} = x)\, dx / \int_{\mathcal{D}_i} K_h(x)\, dx} \to 1,$$

where j_x represents the index such that $D_{ij_x} = x$. By Taylor expansion of $f^*_{j_x}(t \mid D_{ij_x} = x)$ at $x = 0$, we have,

$$\begin{aligned}
\frac{\int_{\mathcal{D}_i} K_h(x) f^*_{j_x}(t \mid D_{ij_x} = x)\,dx}{\int_{\mathcal{D}_i} K_h(x)\,dx} &= \frac{\int_{\mathcal{D}_i} K_h(x)\big[f^*_i(t) + (f^*_i)'(t)x + \frac{(f^*_i)''(t)}{2}x^2\big]\,dx}{\int_{\mathcal{D}_i} K_h(x)\,dx} + O(h^2) \\
&= f^*_i(t) + \frac{\int_{\mathcal{D}_i} K_h(x)\big[(f^*_i)'(t)x + \frac{(f^*_i)''(t)}{2}x^2\big]\,dx}{\int_{\mathcal{D}_i} K_h(x)\,dx} + O(h^2).
\end{aligned}$$

Under assumption (A8) and the condition that $\mathcal{D}_i$ is finite, we have that for some constant $c > 0$,

$$\begin{aligned}
&[\mathbb{E}(f_i(t) \mid \boldsymbol{D}) - f^*_i(t)]^2 \\
&\leq \left(c \int_{\mathcal{D}_i} |x| K_h(x)\,dx \Big/ \int_{\mathcal{D}_i} K_h(x)\,dx + c \int_{\mathcal{D}_i} x^2 K_h(x) \Big/ \int_{\mathcal{D}_i} K_h(x)\,dx\right)^2 + o(1) \\
&\to 0, \qquad \text{as h} \to 0.
\end{aligned}$$

Now for the variance term, by Assumption (A9), we have

$$\operatorname{Var}\left(\sum_{j \neq i} [K_h(D_{ij}) K_h(t_j - t)] \mid \boldsymbol{D}\right) \leq c' \sum_{j \neq i} \left[K_h^2(D_{ij}) \operatorname{Var}(K_h(t_j - t) \mid \boldsymbol{D}_j)\right]$$

Hence, as $h \gg m^{-1/2}$, by Lemma 1, Assumption (2.4) and the fact that $K(\cdot)$ is positive and bounded, we take Taylor expansion again and obtain that

$$\begin{aligned}
\mathrm{Var}(f_i(t) \mid \boldsymbol{D}) &\leq c' m^{-1} \frac{\int_{\mathcal{D}_i} K_h^2(x)\big[f_{j_x}^*(t \mid D_{ij_x} = x)\big(R(K)/h - f_{j_x}^*(t \mid D_{ij_x} = x)\big)\big]\,dx + O(1)}{(\int_{\mathcal{D}_i} K_h(x)\,dx)^2} \\
&\leq c' m^{-1} \frac{\int_{\mathcal{D}_i} \frac{R(K)}{h} K_h^2(x)\big[f_i^*(t) + (f_i^*)'(t)x + \frac{(f_i^*)''(t)}{2}x^2\big]\,dx + O(1)}{(\int_{\mathcal{D}_i} K_h(x)\,dx)^2} \\
&\leq c' m^{-1} \frac{\int_{\mathcal{D}_i} \frac{R(K)}{h} K_h^2(x) f_i^*(t)\,dx + O(h^{-1})}{(\int_{\mathcal{D}_i} K_h(x)\,dx)^2} \\
&\leq c'' m^{-1} \frac{\int_{\mathcal{D}_i} \frac{R(K)}{h} K_h^2(x)\,dx + O(h^{-1})}{(\int_{\mathcal{D}_i} K_h(x)\,dx)^2} \\
&\leq c'' m^{-1} h^{-2} R(K) \frac{\int_{\mathbb{R}} K^2(y)\,dy}{(\int_{\mathcal{D}_i} K_h(x)\,dx)^2} \\
&\leq c''' m^{-1} h^{-2} = o(1),
\end{aligned}$$

for some constant $c'', c''' > 0$. Hence, as $\boldsymbol{D}_i \to \mathcal{D}_i$, combining the bias term and variance term, the consistency result is proved. □

S8.3 Proof of Corollary 1

Proof. Recall that

$$L_i = \frac{(1-\pi_i) f_0(T_i)}{f_i(T_i)}.$$

Then based on the consistency results on π_i and $f_i(t)$ in Propositions 1 (with $\mathcal{N}_i = \{j \in [m], j \neq i\}$ and $h \gg m^{-1}$) and 1, together with the condition that $1/f_i^*(t)$ is bounded and $\pi_i^\tau \leq 1 - \xi$, we have, uniformly for all $i \in [m]$,

$$L_i = (1 + o_{\mathbb{P}}(1)) L_i^*.$$

Then by the condition that $\frac{1}{k^*}\sum_{i=1}^{k^*} L^*_{(i)} \leq \alpha_1 < \alpha < \alpha_2 \leq \frac{1}{k^*+1}\sum_{i=1}^{k^*+1} L^*_{(i)}$ with probability tending to 1 and $\mathrm{Var}\left(L^*_{(k^*)}\right) = o(1)$, it yields that

$$L_{(k)} = L^*_{(k^*)} + o_{\mathbb{P}}(1) = \mathbb{E}\Big\{L^*_{(k^*)}\Big\} + o_{\mathbb{P}}(1).$$

Then based on the definitions of $g^{-1}_{i,+}$ and $g^{-1}_{i,-}$, we have that

$$t_i^{*,+} = g^{-1}_{i,+}\left[\mathbb{E}\left\{L^*_{(k^*)}\right\}\right] \text{ and } t_i^{*,-} = g^{-1}_{i,-}\left[\mathbb{E}\left\{L^*_{(k^*)}\right\}\right],$$

and that

$$t_i^{+} = g^{-1}_{i,+}\left[(1+o_{\mathbb{P}}(1))L_{(k)}\right] \text{ and } t_i^{-} = g^{-1}_{i,-}\left[(1+o_{\mathbb{P}}(1))L_{(k)}\right],$$

based on the condition that $\pi_i^\tau \leq 1-\xi$. Then because $g^{-1}_{i,+}$ and $g^{-1}_{i,-}$ have bounded first derivative, we have

$$t_i^{+} = t_i^{*,+} + o_{\mathbb{P}}(1) \text{ and } t_i^{-} = t_i^{*,-} + o_{\mathbb{P}}(1).$$

By Assumption (A8), f_0 is bounded, then we obtain

$$w_i = w_i^* + o_{\mathbb{P}}(1),$$

uniformly for all $i \in [m]$. □

S8.4 Proof of Theorem 1

Proof. The FDP of the oracle procedure at the thresholding level t can be calculated by

$$
\begin{aligned}
\mathrm{FDP}(t) &= \frac{\sum_{i=1}^m \mathbb{I}\{P_i^{w^*} \le t, \theta_i = 0\}}{\max\{\sum_{i=1}^m \mathbb{I}\{P_i^{w^*} \le t\}, 1\}} \\
&= \frac{\sum_{i=1}^m \mathbb{P}(P_i^{w^*} \le t, \theta_i = 0)}{\max\{\sum_{i=1}^m \mathbb{I}\{P_i^{w^*} \le t\}, 1\}} \cdot \frac{\sum_{i=1}^m \mathbb{I}\{P_i^{w^*} \le t, \theta_i = 0\}}{\sum_{i=1}^m \mathbb{P}(P_i^{w^*} \le t, \theta_i = 0)} \\
&= \frac{\sum_{i=1}^m w_i^*(1-\pi_i^*)t}{\max\{\sum_{i=1}^m \mathbb{I}\{P_i^{w^*} \le t\}, 1\}} \cdot \frac{\sum_{i=1}^m \mathbb{I}\{P_i^{w^*} \le t, \theta_i = 0\}}{\sum_{i=1}^m \mathbb{P}(P_i^{w^*} \le t, \theta_i = 0)} \\
&\le \frac{\sum_{i=1}^m w_i^*(1-\pi_i^\tau)t}{\max\{\sum_{i=1}^m \mathbb{I}\{P_i^{w^*} \le t\}, 1\}} \cdot \frac{\sum_{i=1}^m \mathbb{I}\{P_i^{w^*} \le t, \theta_i = 0\}}{\sum_{i=1}^m \mathbb{P}(P_i^{w^*} \le t, \theta_i = 0)}.
\end{aligned}
$$

Then by Steps 2 and 3 in the proofs of Theorem 1 by replacing w_i's with the true w_i^*'s, and together with the proofs of Theorem 2 in Cai et al. (2022), by Assumption (A10), we have

$$
\sup_{0 \le t \le t_m} \left| \frac{\sum_{i=1}^m \mathbb{I}\{Z_i^{w^*} \ge t, \theta_i = 0\} - \sum_{i=1}^m \mathbb{P}(Z_i^{w^*} \ge t, \theta_i = 0)}{\sum_{i=1}^m \mathbb{P}(Z_i^{w^*} \ge t, \theta_i = 0)} \right| \to 0,
$$

in probability. Then the FDP and FDR are controlled and Theorem 1 is proved. □

S8.5 Proof of Theorem 2

Proof. Note that, the FDP of the data-driven procedure at the thresholding level t can be calculated by

$$
\begin{aligned}
FDP(t) &= \frac{\sum_{i=1}^m \mathbb{I}\{P_i^w \le t, \theta_i = 0\}}{\max[\sum_{i=1}^m \mathbb{I}\{P_i^w \le t\}, 1]} \\
&= \frac{\sum_{i=1}^m \mathbb{P}(P_i^w \le t \mid \theta_i = 0, w_i)\mathbb{P}(\theta_i = 0)}{\max[\sum_{i=1}^m \mathbb{I}\{P_i^w \le t\}, 1]} \cdot \frac{\sum_{i=1}^m \mathbb{I}\{P_i^w \le t, \theta_i = 0\}}{\sum_{i=1}^m \mathbb{P}(P_i^w \le t \mid \theta_i = 0, w_i)\mathbb{P}(\theta_i = 0)},
\end{aligned}
$$

Define the event $A = [\{w_i\}_{i=1}^m : w_i = w_i^* + o(1)]$, then based on the result of Corollary 1, we have that $\mathbb{P}(A) \to 1$. Next, we shall focus on the event A. For $\{w_i\}_{i=1}^m \in A$, uniformly for all $i \in [m]$,

$$
\mathbb{P}(P_i^w \le t | \theta_i = 0, w_i, T_i > 0) = w_i t = [1 + o(1)]w_i^* t,
$$

uniformly for all t defined in the range defined in Step 3 of Theorem 1. The same equality holds if we replace the condition $T_i > 0$ by $T_i < 0$ because the oracle quantity w_i^* is fixed given the sign of T_i. Then we have, uniformly for all $i \in [m]$,

$$\begin{aligned}\mathbb{P}(P_i^w \leq t \mid \theta_i = 0, w_i) &= \mathbb{P}(P_i^w \leq t \mid \theta_i = 0, w_i, T_i > 0)\mathbb{P}(T_i > 0 \mid \theta_i = 0, w_i) \\ &\quad + \mathbb{P}(P_i^w \leq t \mid \theta_i = 0, w_i, T_i < 0)\mathbb{P}(T_i < 0 \mid \theta_i = 0, w_i) \\ &= [1 + o(1)]w_i^* t[\mathbb{P}(T_i > 0 \mid \theta_i = 0, w_i) + \mathbb{P}(T_i < 0 \mid \theta_i = 0, w_i)] \\ &= [1 + o(1)]w_i^* t,\end{aligned}$$

which implies that

$$\mathbb{P}(P_i^w \leq t \mid \theta_i = 0, w_i)\mathbb{P}(\theta_i = 0) = [1 + o(1)]w_i^*(1 - \pi_i^*)t \leq [1 + o(1)]w_i^*(1 - \pi_i^\tau)t.$$

Thus, based on the results of Proposition 1 and Corollary 1 and proofs of Theorems 1 and 1, we obtain that the oracle-assisted weight produces a more conservative procedure asymptotically. This concludes the proof of Theorem 2. □

References

Cai, T., T. T. Cai, and A. Zhang (2016). Structured matrix completion with applications to genomic data integration. *J. Am. Statist. Assoc. 111*(514), 621–633.

Cai, T. T., W. Sun, and Y. Xia (2022). LAWS: A Locally Adaptive Weighting and Screening Approach to Spatial Multiple Testing. *J. Am. Statist. Assoc. 117*, 1370–1383.

Genovese, C. R., K. Roeder, and L. Wasserman (2006). False discovery control with p-value weighting. *Biometrika 93*(3), 509–524.

Hu, J. X., H. Zhao, and H. H. Zhou (2010). False discovery rate control with groups. *J. Am. Statist. Assoc. 105*, 1215–1227.

Krusińska, E. (1987). A valuation of state of object based on weighted mahalanobis distance. *Pattern Recognit. 20*(4), 413–418.

Lei, L. and W. Fithian (2018). Adapt: an interactive procedure for multiple testing with side information. *J. R. Stat. Soc. B 80*(4), 649–679.

Li, A. and R. F. Barber (2019). Multiple testing with the structure-adaptive benjamini–hochberg algorithm. *J. R. Stat. Soc. B 81*(1), 45–74.

Lynch, G., W. Guo, S. K. Sarkar, H. Finner, et al. (2017). The control of the false discovery rate in fixed sequence multiple testing. *Electron. J. Stat. 11*(2), 4649–4673.

Medina, I., J. Carbonell, L. Pulido, S. Madeira, S. Götz, A. Conesa, J. Tárraga, A. Pascual-Montano, R. Nogales-Cadenas, J. Santoyo-Lopez, F. García-García, M. Marba, D. Montaner, and J. Dopazo (2010, 07). Babelomics: An integrative platform for the analysis of transcriptomics, proteomics and genomic data with advanced functional profiling. *Nucleic acids research 38*, W210–3.

Ramdas, A. K., R. F. Barber, M. J. Wainwright, and M. I. Jordan (2019). A unified treatment of multiple testing with prior knowledge using the p-filter. *The Annals of Statistics 47*(5), 2790 – 2821.

Scott, D. W. (1992, 8). *Multivariate Density Estimation: Theory, Practice, and Visualization.* Wiley.

Stein, M. L. (1995). Fixed-domain asymptotics for spatial periodograms. *J. Am. Statist. Assoc. 90*(432), 1277–1288.

Xia, Y., T. T. Cai, and W. Sun (2020). GAP: A General Framework for Information Pooling in Two-Sample

Sparse Inference. *J. Am. Statist. Assoc. 115*, 1236–1250.

Yurko, R., M. G'Sell, K. Roeder, and B. Devlin (2020). A selective inference approach for false discovery rate control using multiomics covariates yields insights into disease risk. *Proceedings of the National Academy of Sciences 117*(26), 15028–15035.

Zhou, Z., W. Wang, L.-S. Wang, and N. R. Zhang (2018). Integrative DNA copy number detection and genotyping from sequencing and array-based platforms. *Bioinformatics 34*(14), 2349–2355.