Statistica Sinica Preprint No: SS-2024-0002	
Title	A Locally Adaptive Algorithm for Multiple Testing with Network Structure
Manuscript ID	SS-2024-0002
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202024.0002
Complete List of Authors	Ziyi Liang, T. Tony Cai, Wenguang Sun and Yin Xia
Corresponding Authors	Yin Xia
E-mails	xiayin@fudan.edu.cn
Notice: Accepted author version.	

A Locally Adaptive Algorithm for Multiple Testing with Network Structure

Ziyi Liang[a], T. Tony Cai[b], Wenguang Sun[c], Yin Xia[d]

[a]*Department of Mathematics, University of Southern California*

[b]*Department of Statistics and Data Science, University of Pennsylvania*

[c] *School of Management and Center for Data Science, Zhejiang University*

[d]*Department of Statistics and Data Science, Fudan University*

Abstract

Incorporating auxiliary information alongside primary data can significantly enhance the accuracy of simultaneous inference. However, existing multiple testing methods face challenges in efficiently incorporating complex side information, especially when it differs in dimension or structure from the primary data, such as network side information. This paper introduces a locally adaptive structure learning algorithm (LASLA), a flexible framework designed to integrate a broad range of auxiliary information into the inference process. Although LASLA is specifically motivated by the challenges posed by network-structured data, it also proves highly effective with other types of side information, such as spatial locations and multiple auxiliary sequences.

LASLA employs a p-value weighting approach, leveraging structural insights to derive data-driven weights that prioritize the importance of different hypotheses. Our theoretical analysis demonstrates that LASLA asymptotically controls the false discovery rate (FDR) under independent or weakly dependent p-values, and achieves enhanced power in scenarios where the auxiliary data provides valuable side information. Simulation studies are conducted to evaluate LASLA's numerical performance, and its efficacy is further illustrated through two real-world applications.

Key words and phrases: Covariate-assisted inference, Distance matrix, False discovery rate, p-value weighting, Structure Learning.

1. Introduction

1.1 Motivating application for network data

Statistical analysis of network-structured data is an important topic with a wide range of applications. Our study is motivated by genome-wide association studies (GWAS), where a primary objective is to identify disease-associated single-nucleotide polymorphisms (SNPs) across diverse populations. Previous studies have indicated that linkage analysis can provide insights into the genetic basis of complex diseases. Particularly, SNPs in linkage disequilibrium (LD) can jointly contribute to the representation of the disease phenotype (Schaub et al., 2012; Joiret et al., 2019). However, existing research in GWAS has often overlooked or underutilized the LD network

information, representing a significant limitation. Therefore, it is essential to develop new integrative analytical tools that can effectively combine GWAS data with auxiliary data from linkage analysis.

Let m denote the number of SNPs and $[m] = \{1, \ldots, m\}$. The primary data, as provided by GWAS, consists of a list of test statistics $\boldsymbol{T} = \{T_i : i \in [m]\}$, which assess the association strength of individual SNPs with a phenotype of interest. Let $\boldsymbol{P} = \{P_i : i \in [m]\}$ denote the corresponding p-values. The auxiliary data, as provided by linkage analysis, is a matrix comprising pairwise correlations $\boldsymbol{S} = (r_{ij} : i, j \in [m])$, where r_{ij} measures the LD correlation between SNP i and SNP j. To illustrate, we construct an undirected LD graph based on the pairwise LD correlation in Figure 1, where each node represents an SNP, and an edge is drawn to connect two SNPs if their correlation exceeds a pre-determined cutoff.

Incorporating the network structured auxiliary data, such as the LD correlation matrix, is desirable as it may improve the power and interpretability of analysis. However, developing a principled approach that cross-utilizes data from different sources is a challenging task. Firstly, the primary data $\boldsymbol{T}$ and the auxiliary data $\boldsymbol{S}$ are usually collected from different populations. For instance, in our data analysis detailed in Section 5.1, the target population in GWAS consists of 77,418 individuals with Type 2 diabetes, while the auxiliary data for linkage analysis are collected from the general population [1000 Genomes (1000G) Phase 3 Database]. Secondly, the dimensions of $\boldsymbol{T}$ and $\boldsymbol{S}$ may not match as in the conventional settings. Specifically,

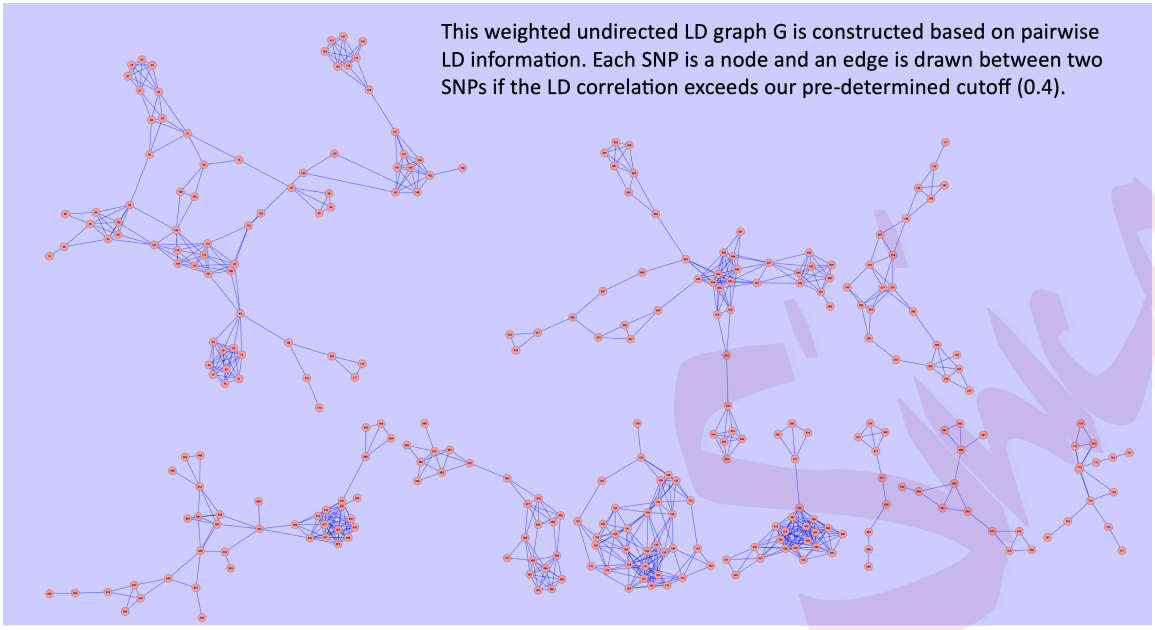


Figure 1: Auxiliary linkage data provide structural information that can be utilized to identify significant SNPs.

$\boldsymbol{T} \in \mathbb{R}^m$ is a vector of statistics while $\boldsymbol{S} \in \mathbb{R}^{m \times m}$ is a network matrix.

Motivated by the observation that SNPs in the same sub-network can exhibit similar distributional characteristics and tend to work together in GWAS (Schaub et al., 2012), this article develops a locally adaptive structure learning algorithm (LASLA) which integrates the structural knowledge in auxiliary data by computing a set of data-driven weights $\{w_i : i \in [m]\}$ to adjust the p-values $\{P_i : i \in [m]\}$. In summary, LASLA follows a two-step strategy:

Step 1: Learn the relational or structural knowledge of the high-dimensional parameter through auxiliary data.

Step 2: Apply the structural knowledge by adaptively placing differential weights w_i on p-values P_i, for $i \in [m]$.

While LASLA is unique in its ability to directly leverage network side information, its versatility extends beyond networks. For instance, when identifying significant regions using functional Magnetic Resonance Imaging (fMRI) data, as discussed in Section 5.2, LASLA can effectively incorporate 3D spatial locations as auxiliary information. Additionally, in high-dimensional regression models aimed at identifying disease-related genetic variants, LASLA can integrate data from related diseases to enhance detection power by prioritizing shared risk factors and genetic variants, as detailed in Section S1 of the Supplementary Material. LASLA offers a flexible and powerful tool that adapts to various forms of auxiliary information, making it suitable for a wide range of applications beyond its original focus on network data integration.

1.2 Related work and our contribution

Structured multiple testing has gained increasing attention in recent years (Cai et al., 2019; Li and Barber, 2019; Castillo and Roquain, 2020; Ren and Candès, 2023). The strategy is to augment the sequence of primary statistics $\boldsymbol{T} = \{T_i : i \in [m]\}$, which may consist of p-values or z-values, with an auxiliary sequence $\boldsymbol{U} = \{U_i : i \in [m]\}$ in order to enhance the statistical power and accuracy of inference. The auxiliary data can be collected through various methods, including: (a) external

sources, such as prior studies and secondary datasets (Ignatiadis et al., 2016; Basu et al., 2018); (b) internal sources within the same dataset, achieved by carefully constructing independent sequences (Cai et al., 2019; Xia et al., 2020); (c) intrinsic patterns associated with the data, such as the natural order of data streams (Foster and Stine, 2008; Lynch et al., 2017) or spatial locations (Lei et al., 2020; Cai et al., 2022). The aforementioned studies primarily focus on an auxiliary sequence $\boldsymbol{U} \in \mathbb{R}^m$ that has the same dimension as $\boldsymbol{T} \in \mathbb{R}^m$. However, there has been limited research conducted on exploiting a wider variety of side information, such as LD matrices or multiple auxiliary sequences.

Moreover, existing structure-adaptive methods (Cai et al., 2019; Li and Barber, 2019; Xia et al., 2020) primarily focus on leveraging sparsity structures in auxiliary data. However, research has shown that other forms of structural information can also significantly enhance inference (Roeder and Wasserman, 2009; Peña et al., 2011; Lei and Fithian, 2018; Fu et al., 2022). For instance, Roeder and Wasserman (2009) illustrates the power gains achieved by constructing weights based on signal amplitudes, Fu et al. (2022) demonstrates the advantages of adjusting for heteroscedasticity, and Peña et al. (2011) highlights the benefits of adapting to hierarchical structures. Despite these advancements, most methods typically concentrate on a specific type of structural information. Therefore, developing a unified framework that can systematically integrate various types of side information is highly desirable.

Our contributions are three-fold. Firstly, LASLA employs a carefully designed

distance matrix to characterize the relational knowledge of the p-values, providing a principled and generic strategy for integrative inference. Secondly, LASLA can leverage various types of structural information by deriving data-driven weights to emulate a hypothetical oracle (Sun and Cai, 2007; Heller and Rosset, 2021), resulting in significant power improvement compared to existing methods (Roquain and Van De Wiel, 2009; Ignatiadis and Huber, 2021; Cai et al., 2022). Lastly, we have developed theory to demonstrate that LASLA asymptotically controls the False Discovery Proportion (FDP) and the False Discovery Rate (FDR) for both independent and weakly dependent p-values, while also achieving proven power gain under mild conditions.

The rest of this article is organized as follows. Section 2 presents the problem formulation and provides details on the development of the LASLA procedure. Section 3 investigates the theoretical properties of LASLA in the independent case and establishes its superiority in ranking. Section 4 presents simulation results comparing LASLA with existing methods. In Section 5, we illustrate LASLA through two real-world applications. Additional simulation results, theoretical results for the dependent case, and proofs of the theories can be found in the Supplementary Material.

2. A Locally Adaptive Structure Learning Algorithm

We present the problem formulation and the basic framework in Section 2.1, followed by the derivation of the oracle rule under a suitable working model in Section 2.2.

Subsequently, we discuss the strategies for approximating the related unknown quantities and derive the oracle-assisted weights in Section 2.3. We address the selection of a threshold for the weighted p-values and present the LASLA procedure in Section 2.4. Finally, we demonstrate the superiority of the LASLA weights over conventional sparsity-adaptive weights in Section 2.5.

2.1 Problem formulation and basic framework

Suppose we are interested in testing m hypotheses:

$$H_{0,i} : \theta_i = 0 \quad \text{vs.} \quad H_{1,i} : \theta_i = 1, \quad i \in [m], \tag{2.1}$$

where $\boldsymbol{\theta} = (\theta_i : i \in [m])$ is a vector of binary random variables indicating the existence or absence of signals at the testing locations. A multiple testing rule can be represented by a binary vector $\boldsymbol{\delta} = (\delta_i : i \in [m]) \in \{0,1\}^m$, where $\delta_i = 1$ indicates that we reject $H_{0,i}$ and $\delta_i = 0$ otherwise. In this paper we focus on problem (2.1) with the goal of controlling the FDP and FDR (Benjamini and Hochberg, 1995) defined respectively by:

$$\mathrm{FDP}(\boldsymbol{\delta}) = \frac{\sum_{i=1}^m (1-\theta_i)\delta_i}{\max(\sum_{i=1}^m \delta_i, 1)}, \quad \text{and FDR} = \mathbb{E}\left\{\mathrm{FDP}(\boldsymbol{\delta})\right\}.$$

An ideal procedure should maximize power, which is interpreted as the expected number of true positives:

$$\mathrm{ETP} = \mathbb{E}\left\{\sum_{i=1}^m \theta_i \delta_i\right\}.$$

Our approach involves the calculation of a distance matrix $\boldsymbol{D} = (1 - r_{ij}^2 : i, j \in$

$[m]$) based on the LD matrix $\boldsymbol{S} = (r_{ij}^2 : i, j \in [m])$, where $r_{ij} \in [0, 1]$ represents the Pearson's correlation coefficient that measures the LD between SNPs i and j. Note that the distance matrix $\boldsymbol{D}$ contains the same information as $\boldsymbol{S}$, as the transformation from $\boldsymbol{S}$ to $\boldsymbol{D}$ involves a simple linear operation: $1 - r_{ij}^2$, applied to each entry in the LD matrix. This transformation facilitates our definition of the local neighborhood and subsequent data-driven algorithms. Specifically, SNPs with higher correlation r_{ij} are considered "closer" in distance. The driving assumption behind our methodological development is that clusters of SNPs with small pairwise distances are part of the same local neighborhood and demonstrate similar distributional characteristics. Conversely, SNPs that are widely separated are more likely to belong to different neighborhoods and exhibit distinct patterns in terms of sparsity levels and distributions. Deriving a distance matrix $\boldsymbol{D}$ from the LD matrix $\boldsymbol{S}$ is a relatively straightforward process. In Section S2 of the Supplementary Material, we discuss extensions that allow for the construction of distance matrices using other types of side information. Harnessing the heterogeneity across different neighborhoods plays a crucial role in constructing informative weights for the p-values. By incorporating weights that encode the network structure, the identification of SNP clusters is facilitated, enabling a more accurate representation of the underlying biological processes and genetic basis of complex diseases. These insights serve as the motivation for our proposed LASLA procedure, which aims to enhance both the interpretability and power of FDR analysis.

2.2 The oracle rule under a working model

One effective strategy to incorporate relevant domain knowledge in multiple testing is through the use of p-value weighting (Genovese et al., 2006; Sun et al., 2015; Basu et al., 2018). The proposed LASLA builds upon this strategy. This subsection introduces an oracle procedure under an ideal setting. In Sections 2.3, we present details for the development of a data-driven algorithm that emulates the oracle. This algorithm utilizes a distance matrix learned from auxiliary data to construct informative weights for the p-values.

Assume the primary test statistics $\{T_i : i \in [m]\}$ follow a hierarchical model:

$$\theta_i \overset{\text{ind}}{\sim} \text{Bernoulli}(\pi_i^*), \quad \mathbb{P}(T_i \leq t|\theta_i) = (1-\theta_i)F_0(t) + \theta_i F_{1i}^*(t), \tag{2.2}$$

where $F_0(t)$ and $F_{1i}^*(t)$ respectively represent the null and alternative cumulative distribution functions (CDF) of T_i. Then T_i obeys the following mixture distribution

$$F_i^*(t) := \mathbb{P}(T_i \leq t) = (1-\pi_i^*)F_0(t) + \pi_i^* F_{1i}^*(t). \tag{2.3}$$

We would like to provide further clarification on our notations. Firstly, Model (2.2) is utilized as a working model to provide an approximation of the true data-generating process. Its primary goal is to facilitate the derivation of our data-driven weights. Secondly, the incorporation of π_i^* and $F_{1i}^*(t)$ aims to account for the potential heterogeneity across testing units, which arises due to the availability of auxiliary data. In contrast, the null distribution $F_0(t)$ is assumed to be known and remains the same across all testing units.

The optimal testing rule, which has the largest ETP among all valid marginal FDR/FDR procedures (Sun and Cai, 2007; Cai et al., 2019; Heller and Rosset, 2021), is a thresholding rule based on the local false discovery rate (lfdr):

$$L_i^* = \mathbb{P}(\theta_i = 0|T_i) = (1 - \pi_i^*) f_0(T_i)/f_i^*(T_i),$$

where $f_0(\cdot)$ and $f_i^*(\cdot)$ respectively denote the density functions corresponding to the CDFs $F_0(\cdot)$ and $F_i^*(\cdot)$. A step-wise algorithm (e.g. Sun and Cai, 2007) can be used to determine a data-driven threshold along the L_i^* ranking. Let $L_{(1)}^* \leq \cdots \leq L_{(m)}^*$ be the ordered statistics of L_i^*, and $H_{(1)}, \ldots, H_{(m)}$ be the corresponding null hypotheses. At FDR level α, the rejection threshold is determined by $k_* = \max\left\{j : j^{-1}\sum_{i=1}^{j} L_{(i)}^* \leq \alpha\right\}$, and the algorithm rejects $H_{(1)}, \ldots, H_{(k_*)}$.

2.3 Data-driven weights

It is essential to note that π_i^* and $f_i^*(t)$ represent hypothetical and non-accessible oracle quantities that cannot be directly employed in data-driven algorithms. Therefore, we develop data-driven quantities to approximate the oracle quantities.

In our derivation of the data-driven quantity to approximate π_i^*, we utilize the screening strategy outlined in the work of Cai et al. (2022). To ensure completeness, we reiterate the key steps below. Let D_{ij} be the distance between entries i and j. A key assumption is that π_i^* is close to π_j^* if D_{ij} is small (and the distributional informations for the two hypotheses are identical if D_{ij}=0), which enables us to estimate π_i^* by borrowing strength from neighboring points. Let K: $\mathbb{R} \to \mathbb{R}$ be a positive,

bounded, and symmetric kernel function that satisfies the following conditions:

$$\int_{\mathbb{R}} K(x)dx = 1, \quad \int_{\mathbb{R}} xK(x)dx = 0, \quad \int_{\mathbb{R}} x^2K(x)dx = \sigma_K^2 < \infty. \tag{2.4}$$

Define $K_h(x) = h^{-1}K(x/h)$, where h is the bandwidth, and $V_h(i,j) = K_h(D_{ij})/K_h(0)$. We construct a data-driven quantity that resembles π_i^* by

$$\pi_i = 1 - \frac{\sum_{j\neq i}[V_h(i,j)\mathbb{I}\{P_j > \tau\}]}{(1-\tau)\sum_{j\neq i} V_h(i,j)}, \quad i \in [m], \tag{2.5}$$

where τ is a screening parameter that will be chosen in Section S4 of the Supplementary Material. Similarly, we introduce a data-driven quantity that resembles the hypothetical density $f_i^*(t)$ for $i \in [m]$ by:

$$f_i(t) = \frac{\sum_{j\neq i}[V_h(i,j)K_h(T_j - t)]}{\sum_{j\neq i} V_h(i,j)}, \tag{2.6}$$

which describes the likelihood of T_i taking a value in the vicinity of t. Consequently, the data-driven lfdr is given by

$$L_i = (1-\pi_i)f_0(T_i)/f_i(T_i). \tag{2.7}$$

Let $L_{(1)} \leq \ldots \leq L_{(m)}$ denote the sorted values of L_i. Following the thresholding approach in Section 2.2, we choose $L_{(k)}$ to be the rejection threshold, where $k = \max\{j : j^{-1}\sum_{i=1}^{j} L_{(i)} \leq \alpha\}$.

Now we are ready to develop the oracle-assisted weights that mimic thresholding rules based on lfdr given in (2.7). Without loss of generality we assume that the function f_0 is symmetric about zero. In cases where this assumption does not hold, we can always transform the primary statistics into z-statistics or t-statistics. Intuitively,

the thresholding rule $L_i < t$ is approximately equivalent to:

$$T_i < t_i^- \text{ or } T_i > t_i^+. \tag{2.8}$$

Here $t_i^- \leq 0$ and $t_i^+ \geq 0$ are coordinate-specific thresholds that lead to asymmetric rejection regions, which are useful for capturing structures of the alternative distribution, such as unequal proportions of positive and negative signals (Sun and Cai, 2007; Li and Barber, 2019; Fu et al., 2022).

The next step is to derive weights that can emulate the rule (2.8). When $T_i \geq 0$, let $t_i^+ = \infty$ if $(1-\pi_i)f_0(t)/f_i(t) > L_{(k)}$ for all $t \geq 0$, else let

$$t_i^+ = \inf\big\{t \geq 0 : (1-\pi_i)f_0(t)/f_i(t) \leq L_{(k)}\big\} \tag{2.9}$$

The rejection rule $T_i > t_i^+$ is equivalent to the p-value rule $P_i < 1 - F_0(t_i^+)$, where P_i is the one-sided p-value. Define the weighted p-values as $\{P_i^w = P_i/w_i : i \in [m]\}$. Intuitively, a larger rejection region $T_i > t_i^+$ implies a proportionally larger weight w_i to prioritize the rejection of H_i. Therefore, if we choose a universal threshold for all P_i^w, using weight $w_i = 1 - F_0(t_i^+)$ can effectively emulate the rule $T_i > t_i^+$.

Similarly, when $T_i < 0$, let $t_i^- = -\infty$ if $(1-\pi_i)f_0(t)/f_i(t) > L_{(k)}$ for all $t \leq 0$, else let

$$t_i^- = \sup\big\{t \leq 0 : (1-\pi_i)f_0(t)/f_i(t) \leq L_{(k)}\big\} \tag{2.10}$$

The corresponding weight is given by $w_i = F_0(t_i^-)$.

To ensure the robustness of the algorithm, we set $w_i = \max\{w_i, \xi\}$ and $w_i = \min\{w_i, 1-\xi\}$, where $0 < \xi < 1$ is a small constant. In all numerical studies, we

set $\xi = 10^{-5}$. To expedite the calculation and facilitate our theoretical analysis, we propose to use only a subset of p-values in the neighborhood to obtain π_i and $f_i(t)$ for $i \in [m]$. Specifically, we define $\mathcal{N}_i = \{j \neq i : D_{ij} \leq a_\epsilon\}$ as a neighborhood set that only contains indices with distance to i smaller than a_ϵ and satisfies $|\mathcal{N}_i| = m^{1-\epsilon}$ for some small constant $\epsilon > 0$. Moreover, we require that $D_{ij_1} \leq D_{ij_2}$ for any $j_1 \in \mathcal{N}_i$, $j_2 \notin \mathcal{N}_i$, and $j_2 \neq i$. This approach has minimal impact on numerical performance, as p-values that are far apart contribute little. We summarize the computation of the data-driven weights in Algorithm 1 below.

Algorithm 1 Data-driven weights

Input: Nominal FDR level α; ϵ specifying the size of the sub-neighborhood; kernel function $K(\cdot)$; primary statistics $\boldsymbol{T} = \{T_i : i \in [m]\}$ and distance matrix $\boldsymbol{D} = (D_{ij} : i, j \in [m])$.
For $i \in [m]$ **do**:
  Estimate π_i as given by (2.5) with summation taken over $\mathcal{N}_i$.
  Estimate $f_i(t)$ as given by (2.6) with summation taken over $\mathcal{N}_i$.
  Compute L_i by (2.7). Denote the sorted statistics by $L_{(1)} \leq \ldots \leq L_{(m)}$.
Let $L_{(k)}$ be the oracle threshold, where $k = \max\{j \in [m] : j^{-1} \sum_{i=1}^{j} L_{(i)} \leq \alpha\}$.
For $i \in [m]$ **do**:
  If $T_i \geq 0$: Compute t_i^+ by (2.9), and the weight by $w_i = 1 - F_0(t_i^+)$.
  If $T_i < 0$: Compute t_i^- by (2.10), and the weight by $w_i = F_0(t_i^-)$.
Output: Data-driven weights $\{w_i : i \in [m]\}$.

2.4 The LASLA procedure

Consider the weighted p-values $\{P_i^w : i \in [m]\}$, where $P_i^w := P_i/w_i$ and $\{w_i : i \in [m]\}$ represent the data-driven LASLA weights. Assume that P_i is independent with w_i, for $i \in [m]$. This assumption only serves as a motivation for the thresholding rule, and

is not required for the theoretical analysis. It follows that, given the set of weights, the expected number of false positives (EFP) for a threshold t^w can be calculated as $t^w \sum_{i=1}^m w_i(1-\pi_i^*)$ under the working model (2.2). Suppose we reject j hypotheses along the ranking provided by $P^w_{(i)}$ for $i \in [m]$, namely, set the threshold as $t^w = P^w_{(j)}$. Then the FDP can be estimated as: $\widehat{\mathrm{FDP}} = \left(P^w_{(j)}/j\right) \sum_{i=1}^m w_i(1-\pi_i)$. We choose the largest possible threshold such that $\widehat{\mathrm{FDP}}$ is less than the nominal level α. Specifically, we find

$$k^w = \max\left\{j : \left(P^w_{(j)}/j\right) \sum_{i=1}^m w_i(1-\pi_i) \le \alpha\right\}, \tag{2.11}$$

and reject $H_{(1)}, \cdots, H_{(k^w)}$, which are the hypotheses with weighted p-values no larger than $P^w_{(k^w)}$.

Remark 1. One could replace (2.11) with other thresholding rules, such as the weighted BH (WBH) procedure (Genovese et al., 2006). However, empirical evidence suggests that WBH tends to be overly conservative. The adaptive WBH method (Ramdas et al., 2019) improves the power of WBH by adjusting to an estimated global sparsity level, but it is less effective when the sparsity level is heterogeneous. See Section S4.6 of the Supplementary Material for a detailed numerical comparison.

2.5 Oracle-assisted weights vs sparsity-adaptive weights

This section presents a comparison between the oracle-assisted weights and the sparsity-adaptive weights discussed in Section S3 of the Supplementary Material. In summary,

the sparsity-adaptive weights only utilize the signal sparsity level π_i^*, whereas the oracle-assisted weights can leverage the structural information embedded in the mixture density $f_i^*(t)$. Consider the oracle sparsity-adaptive weights $w_i^{\text{laws}} = \pi_i^*/(1-\pi_i^*)$ for $i \in [m]$, as reviewed in Section S3. In the case where neighborhood information is provided by spatial locations, the sparsity-adaptive weights are equivalent to the weights employed by the LAWS procedure (Cai et al., 2022). Intuitive examples are provided to illustrate potential information loss associated with the sparsity-adaptive weights (referred to as LAWS weights), and the advantages of the newly proposed weights (referred to as LASLA weights) are highlighted. We focus on scenarios where π_i^* is homogeneous, but it's important to note that the inadequacy of sparsity-adaptive weights extends to cases where π_i^* is heterogeneous. This shortcoming arises from the fact that the sparsity-adaptive weights ignore structural information in alternative distributions.

In the following comparisons, the oracle LASLA weights are obtained via Algorithm 1, with π_i and f_i replaced by their oracle counterparts π_i^* and $f_i^*(t)$. We implement both weights with oracle quantities to emphasize the methodological differences. Consider two examples where $\{T_i : i \in [m]\}$ follow the mixture distribution in (2.3) with $m = 1000$ and $\pi_i^* = 0.1$.

Example 1. Set $F_{1i}^*(t) = \gamma N(3,1) + (1-\gamma)N(-3,1)$, where γ controls the relative proportions of positive and negative signals. We vary γ from 0.5 to 1; when γ approaches 1, the level of asymmetry increases.

Example 2. Set $F_{1i}^*(t) = 0.5N(3, \sigma_i^2) + 0.5N(-3, \sigma_i^2)$, where σ_i controls the shape of the alternative distribution, and $\mathbb{P}(\sigma_i = 1) = 0.5$ and $\mathbb{P}(\sigma_i = \sigma) = 0.5$. We vary σ from 0.2 to 1; the heterogeneity is most pronounced when $\sigma = 0.2$.

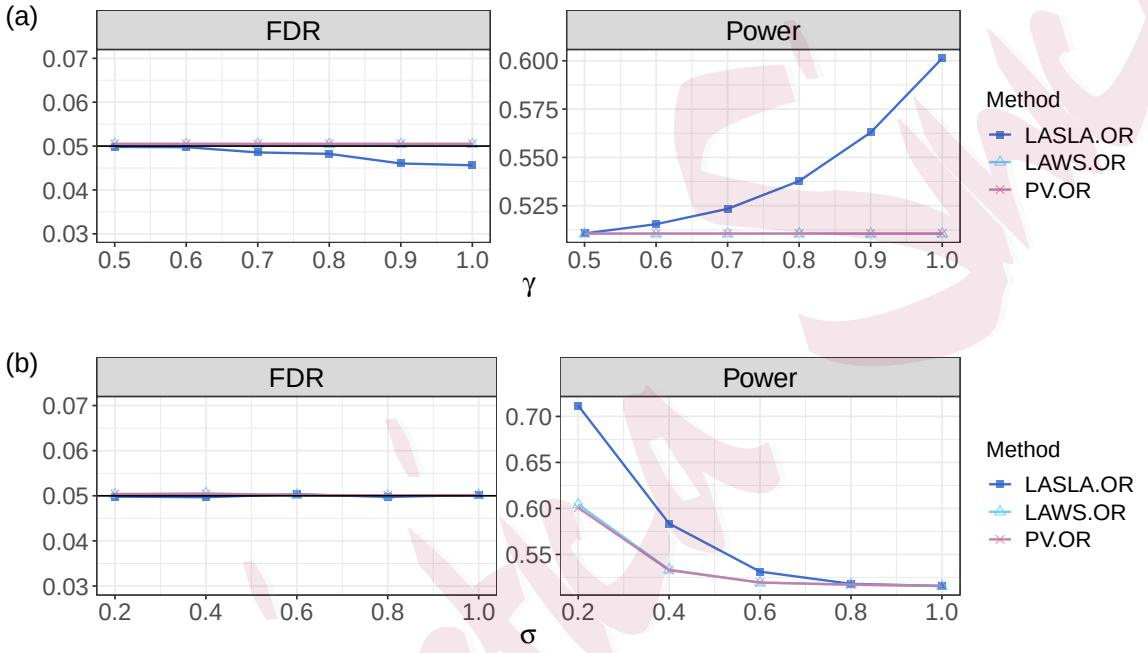


Figure 2: Empirical FDR and power comparison for oracle LASLA (LASLA.OR), oracle LAWS (LAWS.OR) and unweighted oracle p-value procedure (PV.OR) that leverages π_i^* to determine the rejection thresholds, with nominal FDR level $\alpha = 0.05$. Results for LASLA, LAWS and the unweighted oracle procedure are derived by applying the thresholding rule in Section 2.4 respectively with LASLA weights, LAWS weights, and weights equal to 1. (a): Example 1: increasing asymmetry level γ; (b): Example 2: decreasing shape heterogeneity σ.

In both examples, since π_i^* and w_i^{laws} remain constant for all $i \in [m]$, the oracle LAWS reduces to the unweighted p-value procedure (Genovese and Wasserman,

2002). On the contrary, LASLA weights are able to capture the signal asymmetry in Example 1 and the shape heterogeneity in the alternative distribution in Example 2,, resulting in notable power gain in both settings. Additional examples are provided in Section S4.2 to further demonstrate LASLA's strength in leveraging hypothesis-specific information.

3. Theoretical Analysis

We first introduce in Section 3.1 a modified version of the data-driven weights to facilitate the theoretical analysis, and then study in Section 3.2 the FDP and FDR control of LASLA under a marginal independence assumption. Finally, Section 3.3 demonstrates the theoretical power gain of the proposed weighting strategy.

3.1 Modified data-driven weights

To show the asymptotic validity of LASLA with minimal assumptions as shown in Section 3.2, we slightly modify the data-driven weights from Algorithm 1. It is worthy noting that, by imposing stronger conditions as presented in Section S7 of the Supplementary Material, such modification is no longer needed.

Specifically, to facilitate the theoretical analysis, we propose to compute an index-dependent threshold $L^i_{(k)}$ which only relies on $|\mathcal{N}_i|$ statistics as opposed to the universal threshold $L_{(k)}$ defined in Section 2.3. This modification arises from the subtle fact that, though L_i in (2.7) is estimated without using T_i itself (the sign of T_i is

used, but it will not affect the validity of the algorithm), w_i still correlates with T_i since the determination of k in the universal threshold relies on all primary statistics. This intricate dependence structure poses challenges for the theoretical analysis. Therefore, we compute the index-dependent threshold $L^i_{(k)}$ to ensure that given the sign of T_i, w_i is independent of T_i under the null hypothesis at the cost of computational efficiency. This modification has little impact on the numerical performance, but greatly facilitates the establishment of the asymptotic validity of the corresponding weighted multiple testing strategy in Section 3.2. We summarize the modified method in Algorithm 2.

Algorithm 2 Modified data-driven weights

Input: Nominal FDR level α; ϵ specifying the size of the sub-neighborhood; kernel function $K(\cdot)$; primary statistics $\boldsymbol{T} = \{T_i : i \in [m]\}$ and distance matrix $\boldsymbol{D} = (D_{ij} : i, j \in [m])$.
For $i \in [m]$ **do**:
  Estimate π_i by (2.5) and $f_i(t)$ by (2.6) with summation taken over $\mathcal{N}_i$.
  For $j \in \mathcal{N}_i$ **do**:
    Estimate π^i_j by (2.5) and $f^i_j(t)$ by (2.6) with the summation taken over $\mathcal{N}_i$.
    Compute L^i_j as in (2.7) with π_j replaced by π^i_j; $f_j(t)$ by $f^i_j(t)$.
  Denote the sorted statistics by $L^i_{(1)} \leq \ldots \leq L^i_{(|\mathcal{N}_i|)}$.
  Let $k = \max\{j \in [|\mathcal{N}_i|] : j^{-1} \sum_{l=1}^{j} L^i_{(l)} \leq \alpha\}$.
  If $T_i \geq 0$ **then:**
    Compute t_i^+ as given by (2.9), with $L_{(k)}$ replaced by $L^i_{(k)}$;
    Compute the weight by $w_i = 1 - F_0(t_i^+)$.
  If $T_i < 0$ **then:**
    Compute t_i^- as given by (2.10), with $L_{(k)}$ replaced by $L^i_{(k)}$.
    Compute the weight by $w_i = F_0(t_i^-)$.
Output: Modified weights $\{w_i : i \in [m]\}$.

We remark on the distinctions between Algorithm 1 and Algorithm 2. Theo-

retically, the weighted multiple testing procedure based on Algorithm 2 can achieve asymptotic FDR and FDP control with minimal assumptions as shown in Section 3.2, while the validity of the alternative procedure based on Algorithm 1 requires more technical assumptions, as discussed in Section S7 of the Supplementary Material. Practically, weights obtained through both methods are very similar, leading to nearly identical empirical FDR and power performance. Hence, we recommend using Algorithm 1 for practical applications due to its computation efficiency. Note that throughout the paper, the alternative weights from Algorithm 2 are only employed for theoretical purposes in Theorems 1 and 2.

3.2 Asymptotic validity of LASLA

We establish the asymptotic validity of LASLA when each p-value is marginally independent of other p-values under the null. This assumption formally outlined in (A1) provides a comprehensible starting point for our analysis and methodology development, and the dependent cases will be studied in Section S7 of the Supplementary Material.

(A1) P_i is marginally independent of $\boldsymbol{P}_{-i} = \{P_j, j \in [m], j \neq i\}$ under $H_{0,i}$, for $i \in [m]$.

By convention, we assume that the auxiliary variables only affect the alternative distribution but not the null distribution of the primary statistics. Furthermore, we assume that θ_j has no impact on the null distribution of T_i for $i \neq j$, i.e. $\mathbb{P}(T_i \leq t \mid$

$\theta_i = 0, \theta_j) = \mathbb{P}(T_i \leq t \mid \theta_i = 0) = F_0(t)$. This assumption is mild and can be fulfilled by the class of structured probabilistic models (e.g. Goodfellow et al., 2016).

We start with the theoretical analysis on the data-driven estimator π_i, as its consistency is needed for valid FDR control. Let $\boldsymbol{D}_i$ be the ith column of the observed distance matrix $\boldsymbol{D}$, and the corresponding observed network of the indices $i \in [m]$ is considered as a *partial* network. In contrast, we refer to the network as a *full* network when the nodes encompass the entire population under the oracle setting. The associated distance matrix is denoted as $\mathcal{D}$. Define $\mathcal{D}_i$ to be the collection of distances from all nodes in the entire population to index i. Adopting the fixed-domain asymptotics in Stein (1995), we assume that $\boldsymbol{D}_i \to \mathcal{D}_i$ uniformly for all $i \in [m]$ as $m \to \infty$, where $\mathcal{D}_i$ is a continuous[i] finite domain (with respect to coordinate i) in $\mathbb{R}$ with positive measure, and each $d \in \mathcal{D}_i$ is a distance and $0 \in \mathcal{D}_i$. Note that, $\boldsymbol{D}_i$ and $\mathcal{D}_i$ can be viewed as collections of distances respectively measured from partial and full networks with $\boldsymbol{D}_i \subset \mathcal{D}_i$. The details on the theoretical analysis employing fixed-domain asymptotics, i.e., $\boldsymbol{D}_i \to \mathcal{D}_i$ as $m \to \infty$, are relegated to Section S6 in the Supplementary Material. The next assumption requires that, in light of full network information, the distributional quantities vary smoothly in the vicinity of location i.

(A2) For all i, j, $\mathbb{P}(P_j > \tau | \mathcal{D}_j, D_{ij} = x)$ is continuous at x, and has bounded first and second derivatives.

[i]The assumption on the continuity of $\mathcal{D}_i$ can be relaxed. For example, our theory can be easily extended to the case where $\mathcal{D}_i$ is the union of disjoint continuous domains.

It is important to note that marginally independent p-values will become dependent conditional on auxiliary data. The next assumption generalizes the commonly used "weak dependence" notion in Storey (2003). It requires that most of the neighborhood p-values (conditional on auxiliary data) do not exhibit strong pairwise dependence.

(A3) $\text{Var}\left(\sum_{j\in\mathcal{N}_i} K_h(D_{ij})\mathbb{I}\{P_j > \tau\}|\boldsymbol{D}\right) \leq C\sum_{j\in\mathcal{N}_i}\text{Var}\left(K_h(D_{ij})\mathbb{I}\{P_j > \tau\}|\boldsymbol{D}_j\right)$ for some constant $C > 1$, for all $i \in [m]$.

Remark 2. Though marginal independence is assumed, we allow p-values to be weakly dependent conditional on auxiliary data. Additionally, Condition (A3) can be further relaxed by selecting a larger bandwidth h for the kernel $K_h(\cdot)$. For instance, if we choose $h \sim m^{-1/6}$, a common bandwidth selection, then by the proof of Proposition 1, Assumption (A3) can be relaxed to: for some constant $C' > 1$, and for all $i \in [m]$, $\text{Var}\left(\sum_{j\in\mathcal{N}_i} K_h(D_{ij})\mathbb{I}\{P_j > \tau\}|\boldsymbol{D}\right) \leq m^c C'\sum_{j\in\mathcal{N}_i}\text{Var}\left(K_h(D_{ij})\mathbb{I}\{P_j > \tau\}|\boldsymbol{D}_j\right)$ for $c < 5/6$. This relaxation permits the p-values to be highly correlated given the side information.

The next proposition establishes the convergence of π_i (estimated by (2.5) with summation taken over $\mathcal{N}_i$) to an intermediate quantity π_i^τ:

$$\pi_i^\tau = 1 - \frac{\mathbb{P}(P_i > \tau \mid \mathcal{D}_i)}{1-\tau}, \quad 0 < \tau < 1,$$

which offers a good approximation of π_i^* with a suitably chosen τ. Intuitively, when τ is large, null p-values are dominant in the tail area $[\tau, 1]$. Hence $\mathbb{P}(P_i > \tau \mid \mathcal{D}_i)/(1-\tau)$

approximates the overall null proportion.

Proposition 1. *Recall that* $|\mathcal{N}_i| = m^{1-\epsilon}$. *Under Assumptions (A2) and (A3), if* $m^{-1} \ll h \ll m^{-\epsilon}$, *we have, uniformly for all* $i \in [m]$, $\mathbb{E}\left[(\pi_i - \pi_i^{\tau})^2 | \boldsymbol{D}_i\right] \to 0$, *as* $\boldsymbol{D}_i \to \mathcal{D}_i$.

Define $Z_i = \Phi^{-1}(1 - P_i/2)$ and denote by $\boldsymbol{Z} = (Z_1, \ldots, Z_m)^{\intercal}$. We collect below several regularity conditions for proving the asymptotic validity of LASLA.

(A4) Assume that $\sum_{i=1}^m \mathbb{P}(\theta_i = 0 | \mathcal{D}_i) \geq cm$ for some constant $c > 0$ and that $\mathrm{Var}\left[\sum_{i=1}^m I\{\theta_i = 0\} | \mathcal{D}\right] = O(m^{1+\zeta})$ for some $0 \leq \zeta < 1$, where $\mathcal{D} = \{\mathcal{D}_i\}_{i \in [m]}$.

(A5) Define $\mathcal{S}_\rho = \left\{i : 1 \leq i \leq m, |\mu_i| \geq (\log m)^{(1+\rho)/2}\right\}$, where $\mu_i = \mathbb{E}(Z_i)$. For some $\rho > 0$ and $\delta > 0$, $|\mathcal{S}_\rho| \geq [1/(\pi^{1/2}\alpha) + \delta](\log m)^{1/2}$, where $\pi \approx 3.14$ is a constant.

Remark 3. Condition (A4) assumes that the model is sparse and that $\{\theta_i\}_{i=1}^m$ are not perfectly correlated (conditional on auxiliary data). Condition (A5) requires that there are a slowly growing number of signals having magnitude of the order $\{(\log m)/n\}^{(1+\rho)/2}$ if n samples are employed to obtain the p-value.

Let $t^w = P^w_{(k^w)}$, where k^w is calculated based on the step-wise algorithm (2.11) with weights from Algorithm 2. Denote by $\boldsymbol{\delta}^w \equiv \boldsymbol{\delta}^w(t^w) = \{\delta_i^w(t^w) : i \in [m]\}$ the set of decision rules, where $\delta_i^w(t^w) = \mathbb{I}\{P_i^w \leq t^w\}$. Then the FDP and FDR of LASLA are respectively given by

$$\mathrm{FDP}(\boldsymbol{\delta}^w) = \frac{\sum_{i=1}^m (1 - \theta_i)\mathbb{I}\{P_i^w \leq t^w\}}{\max\{\sum_{i=1}^m \mathbb{I}\{P_i^w \leq t^w\}, 1\}}, \text{ and } \mathrm{FDR} = \mathbb{E}\left\{\mathrm{FDP}(\boldsymbol{\delta}^w)\right\}.$$

The next theorem states that LASLA controls both the FDP and FDR at the nominal level asymptotically.

Theorem 1. *Under Assumptions (A1), (A4), (A5) and the conditions in Proposition 1, we have for any* $\varepsilon > 0$*:*

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

In Section S7 of the Supplementary Material, we extend our study to examine the asymptotic control of FDP and FDR for dependent p-values. This analysis is crucial, as it more accurately reflects real-world applications, such as the motivating GWAS example, where complex dependency structures are common. We demonstrate that with additional sufficient regularity conditions on the density functions of the primary statistics, LASLA remains theoretically valid for weakly dependent p-values. Numerical simulations in Section S5 further confirm that LASLA performs robustly across various types of dependencies.

3.3 Asymptotic power analysis

This section provides a theoretical analysis to demonstrate the benefit of the proposed weighting strategy. To simplify the analysis, we assume that the distance matrix $\mathcal{D}$ of the full network is known.

Denote by $\boldsymbol{\delta}^v(t) = \{\delta_i^v(t) : i \in [m]\}$ a class of testing rules based on weighted p-values, where $\delta_i^v(t) = \mathbb{I}\{P_i^v \leq t\}$, $P_i^v = P_i/v_i$ is the weighted p-value, and v_i is the pre-specified weight. It can be shown that (e.g. Proposition 2 of Cai et al. (2022)) under

mild conditions, the FDR of $\boldsymbol{\delta}^v(t)$ can be written as $\mathrm{FDR}\{\boldsymbol{\delta}^v(t)\} = Q^v(t|\mathcal{D}) + o(1)$, where

$$Q^v(t|\mathcal{D}) = \frac{\sum_{i=1}^m (1-\pi_i^*) v_i t}{\sum_{i=1}^m (1-\pi_i^*) v_i t + \sum_{i=1}^m \pi_i^* F_{1i}^*(v_i t|\mathcal{D}_i)}$$

corresponds to the limiting value of the FDR. In what follows, we omit the conditioning on $\mathcal{D}$ for the simplicity of notation. The power of $\boldsymbol{\delta}^v(t)$ can be evaluated using the expected number of true positives with threshold t: $\Psi^v(t) = \sum_{i=1}^m \pi_i^* F_{1i}^*(v_i t)$.

Let $\tilde{w}_i = w_i[\sum_{j=1}^m (1-\pi_j^*)]/[\sum_{j=1}^m (1-\pi_j^*) w_j]$, where w_i's are the modified weights from Algorithm 2. It is easy to see that LASLA and $\boldsymbol{\delta}^{\tilde{w}}(t)$ share the same ranking of hypotheses. The goal is to compare the LASLA weights $(v_i = \tilde{w}_i : i \in [m])$ with the naive weights $\{v_i = 1 : i \in [m]\}$. Denote by $t_o^v = \sup\{t : Q^v(t) \le \alpha\}$ the oracle threshold for the p-values with generic weights $\{v_i : i \in [m]\}$. The oracle procedure with LASLA weights and the (unweighted) oracle p-value procedure (Genovese and Wasserman, 2002) are denoted by $\delta^{\tilde{w}}(t_o^{\tilde{w}})$ and $\delta^1(t_o^1)$, respectively.

Next, we discuss some assumptions needed in our power analysis. The first condition states that weights should be "informative" in the sense that on average, small/large π_i^* correspond to small/large w_i. A similar assumption has been used in Genovese et al. (2006).

(A6) The oracle-assisted weights satisfy

$$\frac{\sum_{i=1}^m (1-\pi_i^*)}{\sum_{i=1}^m (1-\pi_i^*) w_i} \cdot \frac{\sum_{i=1}^m \pi_i^*}{\sum_{i=1}^m \pi_i^* w_i^{-1}} \ge 1.$$

The second condition is concerned with the shape of the alternative p-value dis-

tributions. When the densities are homogeneous, i.e. $F_{1i}^*(t) \equiv F_1^*(t)$, it reduces to the condition that $x \to F_1^*(t/x)$ is a convex function, which is satisfied by commonly used density functions (Hu et al., 2010; Cai et al., 2022).

(A7) For any $0 \le a_i \le 1$, $\min_{i\in[m]} \tilde{w}_i^{-1} \le x_i \le \max_{i\in[m]} \tilde{w}_i^{-1}$ and $t_o^1 / \min_{i\in[m]} \tilde{w}_i^{-1} \le 1$,

$$\sum_{i=1}^{m} a_i F_{1i}^*\left(\frac{t}{x_i}\right) \ge \sum_{i=1}^{m} a_i F_{1i}^*\left(\frac{\sum_{j=1}^m a_j t}{\sum_{j=1}^m a_j x_j}\right).$$

The above two conditions are mild in many practical settings. For example, we checked that both are easily fulfilled in all our simulation studies with the proposed LASLA weights. The next theorem provides insights into why the weighting strategy used in LASLA provides power gain, as we shall see in our numerical studies.

Theorem 2. *Assume Conditions (A1), (A6) and (A7) hold. Then (a)* $Q^{\tilde{w}}(t_o^1) \le Q^1(t_o^1) \le \alpha$*; and (b)* $\Psi^{\tilde{w}}(t_o^{\tilde{w}}) \ge \Psi^{\tilde{w}}(t_o^1) \ge \Psi^1(t_o^1)$.

The theorem implies that (a) if the same threshold t_o^1 is used, then $\delta^{\tilde{w}}(t_o^1)$ has smaller FDR and larger power than $\delta^1(t_o^1)$; (b) the thresholds satisfy $t_o^{\tilde{w}} \ge t_o^1$. Since $\Psi^v(t)$ is non-decreasing in t, we conclude that $\delta^{\tilde{w}}(t_o^{\tilde{w}})$ (oracle procedure with LASLA weights) dominates $\delta^{\tilde{w}}(t_o^1)$ and hence $\delta^1(t_o^1)$ (unweighted oracle p-value procedure) in power.

4. Simulation

This section considers a model that mimics the scenario with network structured side information. Additional simulation results for high-dimensional regression, la-

tent variable model, multiple auxiliary samples, and the implementation details including parameter tuning are relegated to Section S4 of the Supplementary Material. Simulations with dependent data can be found in Section S5. Software implementing the algorithms and replicating all data experiments are available online at `https://github.com/ZiyiLiang/r-Blasla`.

For $i \in [m]$, let $\theta_i \overset{\text{ind}}{\sim} \text{Bernoulli}(0.1)$ denote the existence or absence of the signal at index i. The primary data $\boldsymbol{T} = (T_i : i \in [m])$ are independently generated as $T_i \sim (1-\theta_i)N(0,1)+\theta_i N(\mu_1, 1)$ with μ_1 controlling the signal strength. The distance matrix $\boldsymbol{D} = (D_{ij})_{1\leq i,j\leq m}$ follows $D_{ij} \sim I_{\{\theta_i=\theta_j\}}|N(\mu_2, 0.7)| + I_{\{\theta_i\neq\theta_j\}}|N(1, 0.7)|$, where $0 \leq \mu_2 \leq 1$ controls the informativeness of the distance matrix. Intuitively, if $\theta_i = \theta_j$, then D_{ij} should be relatively small. We investigate two settings.

- Setting 1: Fix $\mu_2 = 0$, $m = 1200$, vary μ_1 from 2.5 to 3 by 0.1;
- Setting 2: Fix $\mu_1 = 3$, $m = 1200$, vary μ_2 from 0 to 1 by 0.2. $\boldsymbol{D}$ becomes less informative as μ_2 gets closer to 1.

Although existing structured multiple testing methods cannot directly incorporate network side information, Yurko et al. (2020) proposes a gradient-boosted trees implementation of the AdaPT procedure (Lei and Fithian, 2018), which indirectly utilizes network information by preprocessing it through hierarchical clustering. For each index i, they generate a vector of indicators denoting cluster memberships, which is then used as side information. In our simulation, we adopt this strategy by cluster-

ing the distance matrix $\boldsymbol{D}$ into 5, 10, and 20 clusters, respectively. We then compare the data-driven LASLA (LASLA.DD) with the clustering-based AdaPT approach and the vanilla BH method, which disregards auxiliary information. The simulation results, averaged over 100 randomized datasets, are summarized in Figure 3.

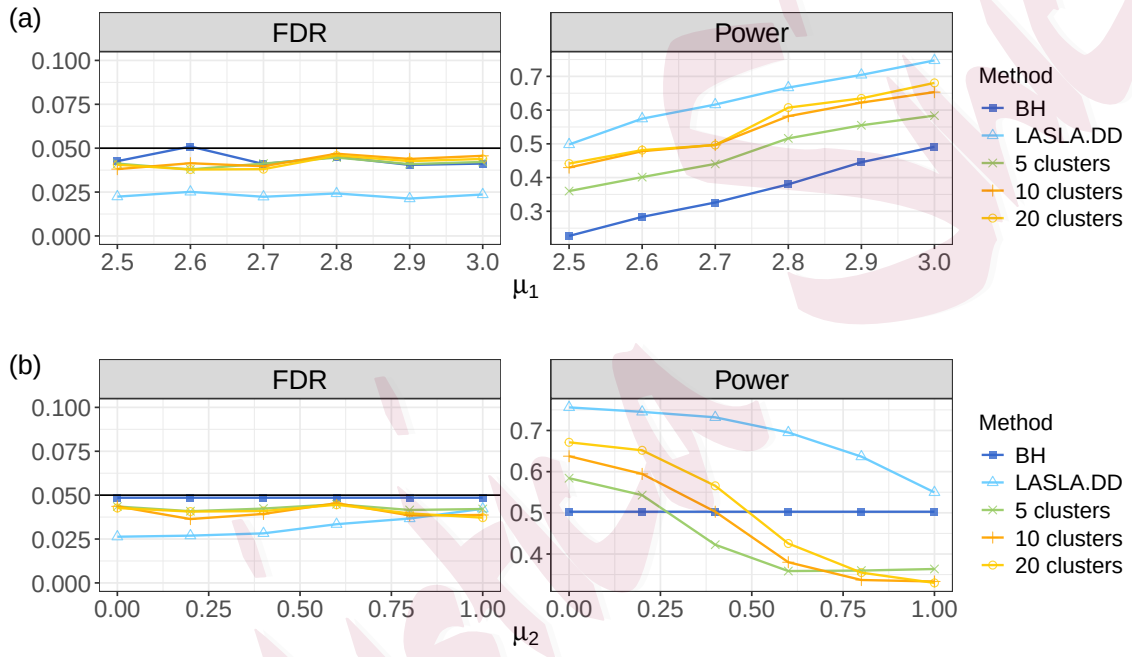


Figure 3: Empirical FDR and power comparison for data-driven LASLA, AdaPT and BH. (a): Setting 1: increasing signal strength μ_1; (b): Setting 2: decreasing informativeness of network side information (controlled by μ_2).

All methods successfully control the FDR at the nominal level. However, in both settings, LASLA demonstrates a substantial power advantage over both the BH and AdaPT methods. In addition, an incorrect choice of the number of clusters for the AdaPT method can lead to either a substantial computational burden or a significant

loss of power.

In Setting 2, we evaluate the performance of LASLA as the informativeness of $\boldsymbol{D}$ varies. As μ_2 approaches 1, $\boldsymbol{D}$ becomes less informative. Notably, even when $\boldsymbol{D}$ becomes completely non-informative ($\mu_2 = 1$), LASLA still outperforms BH. This is because LASLA effectively captures the asymmetry within the alternative distribution of the primary statistics. This finding is consistent with Sun and Cai (2007), which shows that the lfdr (Efron et al., 2001) procedure dominates BH in power. In contrast, the performance of the AdaPT procedure heavily depends on the quality of the side information. This phenomenon is consistently observed in data-sharing high-dimensional regression and latent variable settings, as shown in Figures S4.3-S4.4 in the Supplementary Material.

Furthermore, we extend this network simulation to dependent scenarios in Sections S5.1-S5.2, mimicking the potential dependency structure of GWAS dataset. The results demonstrate that LASLA's performance remains robust across various types of dependencies.

5. Real Data Applications

5.1 Detecting T2D-associated SNPs with auxiliary data from linkage analysis

This section focuses on conducting association studies of Type 2 diabetes (T2D), a prevalent metabolic disease with strong genetic links. Our primary goal is to identify

SNPs associated with T2D in diverse populations. We construct the distance matrix from LD information to gain valuable insights into the genetic basis of complex diseases as previously described in Section 2.1.

Spracklen et al. (2020) performs a meta-analysis to combine 23 studies on a total of 77,418 individuals with T2D and 356,122 controls. For illustration purposes, we randomly choose $m = 5000$ SNPs from Chromosome 6 to be the target of inference. Primary statistics are the z-values provided in Spracklen et al. (2020) and the auxiliary LD matrix is constructed by the genetic analysis tool `Plink` from the 1000 Genomes (1000G) Phase 3 Database. It is important to note that the primary and auxiliary data are collected from different populations and are not matched in dimension.

We apply BH and LASLA at different FDR levels and compare them with the Bonferroni correction which is commonly used in GWAS to control Family-wise Error Rate (FWER). The numbers of rejections by different methods are summarized in Tables 1 and 2. Both BH and LASLA are more powerful than the Bonferroni method. Moreover, at the same FDR level, LASLA makes notably more rejections than BH, and the discrepancy becomes larger as the nominal FDR level increases.

Table 1: Number of rejections by different methods

FDR	0.001	0.01	0.05	0.1
BH	35	61	128	179
LASLA	36	101	184	271

Table 2: Number of rejections by Bonferroni Correction

FWER	0.001	0.01	0.05	0.1
Bonferroni	21	29	35	43

To illustrate the power gain of LASLA over BH, we visualize the rejected hypotheses in Figure 4. Red nodes in the figure present SNPs detected by LASLA but not by BH at FDR level 0.05. Nodes connected by an edge are in linkage disequilibrium. The graph highlights LASLA's ability to leverage the LD matrix's network structure for inference, leading to the identification of clusters of SNPs in LD. In contrast, BH could potentially miss important variants. Notably, LASLA detects T2D-risk variants within the gene CCHCR1, a candidate gene for T2D reported by Brenner (2020).

5.2 Detecting significant brain regions in ADHD patients using fMRI data

In this section, we focus on identifying significant brain regions in patients with Attention Deficit Hyperactivity Disorder (ADHD) using functional Magnetic Resonance Imaging (fMRI) data, incorporating the spatial distance network as side information. The dataset, obtained from the ADHD-200 Sample Initiative, was preprocessed by the Neuro Bureau and is publicly available at `http://neurobureau.projects.nitrc.org/ADHD200/Data.html`.

The dataset includes 931 individuals, comprising 356 diagnosed with ADHD and

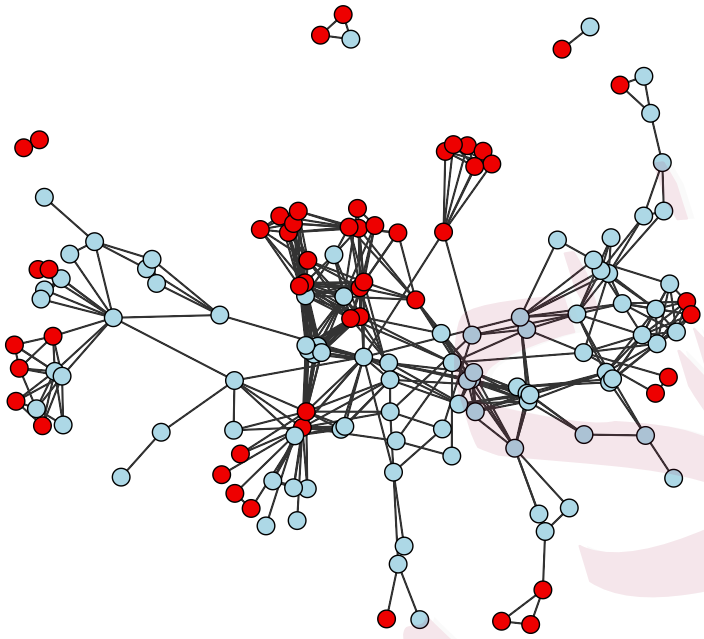
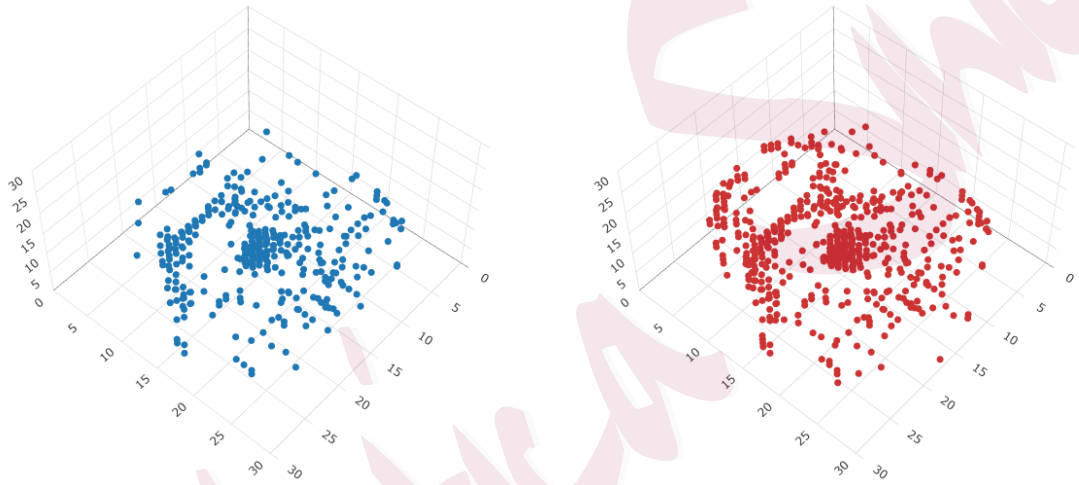


Figure 4: Sub-network identified by LASLA (all nodes). BH only detects the blue nodes.

575 neurotypical controls. Following Li and Zhang (2017), we downsize the original MRI images, with a resolution of $256 \times 198 \times 256$, to a lower resolution of $30 \times 36 \times 30$ by aggregating neighboring pixels into larger blocks. This additional preprocessing step provides two key advantages. First, it enhances the signal-to-noise ratio and reduces potential misalignment issues during measurement, thereby improving the accuracy of the analysis. Second, it effectively reduces dependency by capturing localized correlations within each block.

Next, we perform two-sample t-tests on the aggregated blocks to compare the

two groups, using a normal approximation to calculate the p-values. The results are summarized in Figure 5.2. LASLA reveals a clearer pattern of brain regions that exhibit significant differences between individuals with ADHD and those without. Specifically, the BH procedure identified 349 regions, while LASLA detected 540 at an FDR level of 0.05.



(a) BH rejection regions (b) LASLA rejection regions

Figure 5: Differential brain regions between ADHD patients and controls identified by BH and LASLA at an FDR level of 0.05.

Supplementary Material

The Supplementary material includes numerical implementation details, additional simulations and applications of LASLA, as well as theoretical results for the dependent case and the proofs of all theories.

Acknowledgments

The research of Yin Xia was supported in part by the National Natural Science Foundation of China (Grant No. 12331009, 12022103). The research of Tony Cai was supported in part by the National Science Foundation (Grant DMS-2413106) and the National Institutes of Health (Grants R01-GM129781 and R01-GM123056).

References

Basu, P., T. T. Cai, K. Das, and W. Sun (2018). Weighted false discovery rate control in large-scale multiple testing. *J. Am. Statist. Assoc. 113*(523), 1172–1183.

Benjamini, Y. and Y. Hochberg (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. Roy. Statist. Soc. B 57*(1), 289–300.

Brenner, L. N. e. a. (2020). Analysis of glucocorticoid-related genes reveal cchcr1 as a new candidate gene for type 2 diabetes. *J. Endocr. Soc. 4*(11), bvaa121.

Cai, T. T., W. Sun, and W. Wang (2019). CARS: Covariate assisted ranking and screening for large-scale two-sample inference (with discussion). *J. Roy. Statist. Soc. B 81*, 187–234.

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.

Castillo, I. and É. Roquain (2020). On spike and slab empirical bayes multiple testing. *Ann. Statist. 48*(5), 2548–2574.

Efron, B., R. Tibshirani, J. D. Storey, and V. Tusher (2001). Empirical Bayes analysis of a microarray

experiment. *J. Amer. Statist. Assoc. 96*, 1151–1160.

Foster, D. P. and R. A. Stine (2008). α-investing: a procedure for sequential control of expected false discoveries. *J. R. Stat. Soc. B 70*(2), 429–444.

Fu, L., B. Gang, G. M. James, and W. Sun (2022). Heteroscedasticity-adjusted ranking and thresholding for large-scale multiple testing. *J. Am. Statist. Assoc. 117*(538), 1028–1040.

Genovese, C. and L. Wasserman (2002). Operating characteristics and extensions of the false discovery rate procedure. *J. R. Stat. Soc. B* ***64***, 499–517.

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

Goodfellow, I., Y. Bengio, and A. Courville (2016). *Deep Learning*. MIT Press. `http://www.deeplearningbook.org`.

Heller, R. and S. Rosset (2021). Optimal control of false discovery criteria in the two-group model. *J. Roy. Statist. Soc. B 83*(1), 133–155.

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

Ignatiadis, N. and W. Huber (2021). Covariate powered cross-weighted multiple testing. *J. Roy. Statist. Soc. B 83*(4), 720–751.

Ignatiadis, N., B. Klaus, J. B. Zaugg, and W. Huber (2016). Data-driven hypothesis weighting increases detection power in genome-scale multiple testing. *Nat. Methods 13*(7), 577.

Joiret, M., J. M. Mahachie John, E. S. Gusareva, and K. Van Steen (2019). Confounding of linkage

disequilibrium patterns in large scale dna based gene-gene interaction studies. *BioData Min. 12*(1), 11.

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.

Lei, L., A. Ramdas, and W. Fithian (2020). A general interactive framework for false discovery rate control under structural constraints. *Biometrika 108*(2), 253–267.

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.

Li, L. and X. Zhang (2017). Parsimonious tensor response regression. *Journal of the American Statistical Association 112*(519), 1131–1146.

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.

Peña, E. A., J. D. Habiger, and W. Wu (2011). Power-enhanced multiple decision functions controlling family-wise error and false discovery rates. *Ann. Statist. 39*(1), 556.

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.

Ren, Z. and E. Candès (2023). Knockoffs with side information. *Ann. Appl. Stat. 17*(2), 1152–1174.

Roeder, K. and L. Wasserman (2009). Genome-wide significance levels and weighted hypothesis testing. *Statistical science: a review journal of the Institute of Mathematical Statistics 24*(4), 398.

Roquain, E. and M. A. Van De Wiel (2009). Optimal weighting for false discovery rate control. *Electron.*

J. Stat. 3, 678–711.

Schaub, M. A., A. P. Boyle, A. Kundaje, S. Batzoglou, and M. Snyder (2012). Linking disease associations with regulatory information in the human genome. *Genome Res. 22*(9), 1748–1759.

Spracklen, C., M. Horikoshi, and Y. e. a. Kim (2020). Identification of type 2 diabetes loci in 433,540 east asian individuals. *Nature 582*, 240–245.

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

Storey, J. D. (2003). The positive false discovery rate: a Bayesian interpretation and the q-value. *Ann. Statist. 31*, 2013–2035.

Sun, W. and T. T. Cai (2007). Oracle and adaptive compound decision rules for false discovery rate control. *J. Amer. Statist. Assoc.* ***102***, 901–912.

Sun, W., B. J. Reich, T. T. Cai, M. Guindani, and A. Schwartzman (2015). False discovery control in large-scale spatial multiple testing. *J. R. Stat. Soc. B 77*(1), 59–83.

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.

Ziyi Liang

E-mail: (ziyilian@usc.edu)

T. Tony Cai

E-mail: (tcai@wharton.upenn.edu)

Wenguang Sun

E-mail: (wgsun@zju.edu.cn)

Yin Xia

E-mail: (xiayin@fudan.edu.cn)