Detecting strong signals in gene perturbation experiments: An adaptive approach with power guarantee and FDR control

In this supplement contains proofs, descriptions of methods and simulations not included in the main context. Section A contains proofs to Theorems 2.5, 2.7, 2.8 and 2.10. In section C, we give detailed description of the truncated MLE method and the central moment matching(CM) method. Simulations comparing the ITEB method with them are given in section D. In section E, we provide more details on how the knock-down experiment data is generated.

A Proof of Theorem 2.5, Theorem 2.7, Theorem 2.10 and Theorem 2.8

Let $\Delta_1 := \sqrt{\frac{\log N}{N}}(\tau^2+C)$, and let $t_l := \max(\frac{\log^2 N}{N}, \min(2\alpha_2, \frac{l}{N}))$, $\Delta_{2,l} := 3(\tau^2+C)t_l \log\frac{1}{t_l}$, $\tau_l^2 := [\tau^2 - \Delta_1 - \Delta_{2,l}]_+$ for all $l = 0, 1, \ldots, |A_1|$. Let the oracle estimator be defined as $\tau_*^2 = \frac{\sum_{i\in A_0} x_i^2 - \hat{\sigma}_i^2}{|A_0|}$. Let $p_{i,l} = \tilde{F}(\frac{x_i^2}{\tau_l^2+\hat{\sigma}_i^2})$ be the p-values calculated using τ_l^2 and let $p_{(i),l}$ be the ordered null p-values from small to large. Let $B_{1,l} = \{i \in A_0 : p_{i,l} \leq p_{(s_0),l}\}$, where $s_0 = \max\{s : p_{(s),l} \leq \frac{(s+l)\alpha_1}{N}\}$, and $B_{2,l} = \{i \in A_0 : p_{i,l} \leq \alpha_2\}$. Lemma A.1 contains the deterministic relationships we will use later.

Lemma A.1. *Suppose* $0 < \alpha_1 < \frac{1}{2e}$ *to be fixed and* $\alpha_2 \to 0$ *at a slow rate* ($\alpha_2 \frac{N}{\log^2 N}$ *is bounded away from 0). Under Assumption 2.1, the following statements hold:*

(1)$\lambda(\frac{l\alpha_1}{N}) \lesssim (\frac{N}{l})^{\frac{2}{5}}$, $\frac{\Delta_1}{\tau^2+1}\lambda(\frac{\alpha_1}{N}) \to 0$, $\sup_{l\geq 1}\frac{\Delta_{2,l}}{\tau^2+1}\lambda(\frac{l\alpha_1}{N}) \to 0$, $\sup_{l\geq 0}\frac{\Delta_{2,l}}{\tau^2+1}\lambda(\alpha_2) \to 0$

(2)$\lim_{N\to\infty}\sup_{i\in A_0}\sup_{l\geq 0}\sup_{\alpha\geq\min\left(\alpha_2, \frac{(l\vee 1)\alpha_1}{N}\right)}\frac{P(p_{i,l}\leq\alpha)}{\alpha} = 1$

(3)The estimate $\hat{\tau}^2_{S_k}$ *is non-increasing in the iteration number* k *in the ITEB procedure.*

Proof of A.1 is deferred to section B.

Proof of Theorem 2.5

Proof. Let $E_{k,l} := \{\hat{\tau}^2_{S_k} \geq \tau_l^2\}$. Because $\cup_{l=0}^{|A_1|}\{R_K = l\}$ is a partition of the full space, to prove the statement, we show the following:

$$P(\cup_{l=0}^{|A_1|}\{R_K = l, E^c_{K,l}\}) \leq P(\cup_{l=0}^{|A_1|}\{R_K \leq l, E^c_{K,l}\}) \to 0$$

Define $R_k := |J_k \cap A_1|$ for each iteration k. We also let $R_0 = 0$, and $R_{k-1} \leq R_k$ as a consequence of Lemma A.1, part (3). We prove it by showing that the events $\{R_k \leq l, E^c_{k,l}, E_{k-1,l}\}$ and $\{E^c_{0,l}\}$ do not happen on a properly chosen event M which itself happens with probability approaching 1. The existence of such M is sufficient for our argument:

$$\begin{aligned}\sum_{l=0}^{|A_1|} P(\{R_K \leq l, E^c_{K,l}, M\}) &\leq \sum_{l=0}^{|A_1|} P(R_K \leq l, E^c_{K,l}, E_{K-1,l}, M) + \sum_{l=0}^{|A_1|} P(R_K \leq l, E^c_{K-1,l}, M) \\ &\leq \sum_{l=0}^{|A_1|} P(R_{K-1} \leq l, E^c_{K-1,l}, M) \leq \ldots \leq \sum_{l=0}^{|A_1|} P(E^c_{0,l}, M) = 0\end{aligned}$$

Then $P(\cup_{l=0}^{|A_1|}\{R_K \leq l, E^c_{K,l}\}) \leq P(\cup_{l=0}^{|A_1|}\{R_K \leq l, E^c_{K,l}, M\}) + P(M^c) \to 0$. We now find M which contradicts $\{R_k \leq l, E^c_{k,l}, E_{k-1,l}\}$ and $\{E^c_{0,l}\}$. Let $S^0_k = A_0 \cap S_k$ and $S^1_k = A_1 \cap S_k$ be the set of nulls and non-nulls remaining at iteration k. The relationship below always holds:

$$\hat{\tau}^2_{S_k} \geq \frac{[|A_0|\tau^2_* - \sum_{i \in A_0 \setminus S^0_k} x_i^2 + \sum_{i \in S^1_k}(x_i^2 - \hat{\sigma}_i^2)]_+}{|S^0_k| + |S^1_k|} \tag{1}$$

In other words,

$$\{E^c_{k,l}, E_{k-1,l}, R_k \leq l\} \subseteq \{\frac{[|A_0|\tau^2_* - \sum_{i \in A_0 \setminus S^0_k} x_i^2 + \sum_{i \in S^1_k}(x_i^2 - \hat{\sigma}_i^2)]_+}{|S^0_k| + |S^1_k|} < \tau^2_l, E_{k-1,l}, R_k \leq l\}$$

If $\tau^2_l \leq 0$, the above event will never happen, hence,

$$\begin{aligned}&\{E^c_{k,l}, E_{k-1,l}, R_k \leq l\} \\ \subseteq&\{|A_0|\tau^2_* - \sum_{i \in A_0 \setminus S^0_k}(x_i^2 - \tau^2_l) + \sum_{i \in S^1_k}(x_i^2 - \hat{\sigma}_i^2 - \tau^2_l) < \tau^2_l|A_0|, \tau^2_l > 0, E_{k-1,l}, R_k \leq l\}\end{aligned}$$

When $E_{k-1,l}$ happens and when $R_k \leq l$, the removed nulls $A_0 \setminus S^0_k$ must be in the set $B_l := B_{1,l} \cap B_{2,l}$ for the following reasons. $A_0 \setminus S^0_k \subseteq B_{2,l}$ by definition. Now, we let $p_* := \frac{|J^1_k|\alpha_1}{N}$ be the cut-off for the rejected p-values for the set J^1_k. There are two possibilities : $p_* \leq \alpha_2$ or $p_* > \alpha_2$. We discuss them separately.

1. If $p_* \leq \alpha_2$, the rejected set from ITEB will be the set J^1_k and J^1_k contains at most l non-null hypothesis. Suppose J^1_k contains exactly s_0 null hypotheses. In order for any null hypothesis i to be rejected, it must satisfy $p_i \leq \frac{(s_0+l)\alpha_1}{N}$, and we know there

are s_0 of them. As a result, we know $s_0 = |\{i \in A_0 : p_i \leq \frac{(s_0+l)\alpha_1}{N}\}|$. When $E_{k-1,l}$ happens, we have $p_i \geq p_{i,l}$, thus $s_0 \leq |\{i \in A_0 : p_{i,l} \leq \frac{(s_0+l)\alpha_1}{N}\}|$ or $p_{(s_0),l} \leq \frac{(s_0+l)\alpha_1}{N}$. As a result, we have $s_0 \leq \arg\max\{s : p_{(s),l} \leq \frac{s+l}{N}\alpha_1\}$ and $A_0 \setminus S_k^0 \subseteq B_{1,l}$. Hence, $A_0 \setminus S_k^0 \subseteq B_l$.

2. If $p_* > \alpha_2$, the rejected set from ITEB will be the set J_k^2. In this case, we can show that $B_{2,l} \subseteq B_{1,l}$. Because $p_{i,l} \leq p_i$, everything in J_k^1 will again be rejected if we replace p_i with $p_{i,l}$, in other words, $|B_{1,l}| \geq |J_k^1|$. As a result, p_{**}, the new cut-off p-value for $B_{1,l}$, will be larger than α_2: $p_{**} \geq \frac{|J_k^1|\alpha_1}{N} > \alpha_2$, which in turns lead to the fact that $B_{2,l} \subseteq B_{1,l}$ and $B_l = B_{2,l}$. Hence, $A_0 \setminus S_k^0 \subseteq B_l$.

Let $\tilde{A}_{1,l} := \{i \in A_1 : p_{i,l} \geq \alpha_2\}$. When $E_{k-1,l}$ holds, $\tilde{A}_{1,l} \subseteq S_k^1$, and for any $i \in B_l$ or $i \in A_1 \setminus \tilde{A}_{1,l}$, we have $x_i^2 \geq \tilde{F}_i(\alpha_2)(\tau_l^2 + \hat{\sigma}_i^2)$, thus $(x_i^2 - \hat{\sigma}_i^2) > \tau_l^2$, and

$$\{E_{k,l}^c, E_{k-1,l}, R_k \leq l\} \subseteq \{|A_0|\tau_*^2 - \sum_{i \in B_l}(x_i^2 - \tau_l^2) + \sum_{i \in \tilde{A}_{1,l}}(x_i^2 - \hat{\sigma}_i^2 - \tau_l^2) < \tau_l^2|A_0|, \tau_l^2 > 0\}$$

We can construct M based on the equation above. We define the following six events:

$$M_1 = \{\forall l = 0, 1, \ldots, |A_1| : |B_{1,l}| \leq \frac{(\log^2 N \vee l)}{N}|A_0|\}$$
$$M_2 = \{\forall l = 0, 1, \ldots, |A_1| : |B_{2,l}| \leq 2\alpha_2|A_0|\}$$
$$M_3 = \{\forall l = 0, 1, \ldots, |A_1| : (\max_{A_\alpha \subseteq A_0 : |A_\alpha| \leq t_l|A_0|} \sum_{i \in A_\alpha} x_i^2) \leq 2.5(\tau^2 + C)t_l|A_0|\log\frac{1}{t_l}\}$$
$$M_4 = \{\forall l = 0, 1, \ldots, |A_1| : \sum_{i \in \tilde{A}_{1,l}}(x_i^2 - \hat{\sigma}_i^2 - \tau_l^2) \geq -(\tau^2 + 1)(1 - \gamma)\sqrt{|A_1|\log N}\}$$
$$M_5 = \{\sum_{i \in A_1}(x_i^2 - \hat{\sigma}_i^2 - \tau^2) \geq -(\tau^2 + 1)(1 - \gamma)\sqrt{|A_1|\log N}\}$$
$$M_0 = \{|\tau^2 - \tau_*^2| \leq (1 - \sqrt{\gamma})\Delta_1\}$$

Let $M = \cap_{j=0}^5 M_j$. Lemma A.2 states that M happens with probability approaching 1, whose proof is deferred to section B.

Lemma A.2. *Under Assumption 2.1 and Assumption 2.2, with $\alpha_1 < \frac{1}{2e}$ being a positive constant and $\alpha_2 \to 0$ at a slow rate ($\alpha_2 \frac{N}{\log^2 N}$ is bounded away from 0), M_0, M_1, M_2, M_3, M_4 and M_5 happen with probability approaching 1.*

When M_1 and M_2 hold, we have $|B_l| = |B_{l,1} \cap B_{2,l}| \leq t_l|A_0|$, and if M_3 also holds, we have

$$\sum_{i \in B_l} x_i^2 \leq \max_{A_\alpha \in A_0, |A_\alpha| \leq t_l |A_0|} \sum_{i \in A_\alpha} x_i^2 \leq 2.5(\tau^2 + C)|A_0| t_l \log(\frac{1}{t_l}) \leq \frac{5}{6}|A_0|\Delta_{2,l}$$

When M_5 holds, we have $\sum_{i \in \tilde{A}_{1,l}}(x_i^2 - \hat{\sigma}_i^2 - \tau_l^2) \geq -\sqrt{\gamma}|A_0|\Delta_1$. Therefore, the following is true,

$$\begin{aligned}
\{E_k^c, E_{k-1}, R_k \leq l, M\} &\subseteq \{|A_0|(\tau^2 - (1-\sqrt{\gamma})\Delta_1) - \frac{5}{6}|A_0|\Delta_{2,l} - \sqrt{\gamma}|A_0|\Delta_1 < |A_0|\tau_l^2, \tau_l^2 > 0\} \\
&= \{(\tau^2 - (1-\sqrt{\gamma})\Delta_1) - \frac{5}{6}\Delta_{2,l} - \sqrt{\gamma}\Delta_1 < (\tau^2 - \Delta_1 - \Delta_{2,l})\} \\
&= \{\frac{5}{6}\Delta_{2,l} > \Delta_{2,l}\} = \emptyset
\end{aligned}$$

In the step 0, we use all points to estimate $\hat{\tau}_{S_0}^2 = \frac{|A_0|\tau_*^2 + \sum_{i \in A_1}(x_i^2 - \hat{\sigma}_i^2)}{N}$. When M_0 holds, we have $|A_0|\tau_*^2 \geq |A_0|(\tau^2 - (1-\sqrt{\gamma})\Delta_1)$; when M_5 holds, we have $\sum_{i \in A_1}(x_i^2 - \hat{\sigma}_i^2 - \tau^2) \geq -|A_0|\sqrt{\gamma}\Delta_1$, thus we have $\{E_{0,l}^c, M\} = \{\hat{\tau}_{S_0}^2 < [\tau^2 - \Delta_1 - \Delta_{2,0}]_+, M\} = \emptyset$. $\square$

Remark A.3. *When $\delta > 0$, we need to replace $\hat{\sigma}_i^2$ with $(1+\delta)\hat{\sigma}_i^2$, and correspondingly, σ_i^2 with $(1+\delta)\sigma_i^2$ at several places. For example, we will modify equation (1) into the following*

$$\hat{\tau}_{S_k}^2 \geq \frac{[|A_0|\tau_*^2 - \sum_{i \in A_0 \setminus S_k^0} x_i^2 + \sum_{i \in S_k^1}(x_i^2 - (1+\delta)\hat{\sigma}_i^2)]_+}{|S_k^0| + |S_k^1|}$$

However, this will not change our final results when δ is of order $O(1/\sqrt{N})$, hence, we leave out δ in our analysis for convenience.

Proof of Theorem 2.7

Proof. As $\hat{\tau}_{S_k}^2$ is non-increasing, and in order for a point to be removed at any iteration k, it must be greater than $\hat{\tau}_{S_K}^2$: $x_i^2 \geq \tilde{F}_i^{-1}(\alpha_2)(\hat{\tau}_{S_k}^2 + \hat{\sigma}_i^2) \Rightarrow x_i^2 - \hat{\sigma}_i^2 > \hat{\tau}_{S_k}^2 \geq \hat{\tau}_{S_K}^2$. Let $S_K^0 = S_K \cap A_0$, we have

$$\sum_{i \in S_K^0}(x_i^2 - \hat{\sigma}_i^2) \leq |A_0|\tau_*^2 - (|A_0| - |S_K^0|)\hat{\tau}_{S_K}^2$$

For a point $i \in A_1$, in order for it to not be removed, it need to satisfy the following criterion:

$$x_i^2 \leq \tilde{F}_i^{-1}(\alpha_2)(\hat{\sigma}_i^2 + \hat{\tau}_{S_k}^2) \text{ or } x_i^2 \leq \tilde{F}_i^{-1}(\frac{l_i\alpha_1}{N})(\hat{\sigma}_i^2 + \hat{\tau}_{S_k}^2)$$

where l_i is the order of the p-value of x_i^2. As a result, at the last iteration K, we have

$$\hat{\tau}_{S_K}^2 \leq [\frac{|A_0|\tau_*^2 - (|A_0| - |S_{K_0}|)\hat{\tau}_{S_K}^2}{|S_K|} + \frac{\sum_{i\in A_1} \lambda((\frac{l_i}{N}\alpha_1) \wedge \alpha_2)(\hat{\sigma}_i^2 + \hat{\tau}_{S_K}^2)}{|S_K|}]_+$$

If $\hat{\tau}_{S_K}^2 = 0$, we have proved our statement; otherwise, the term insider the positive operator is positive, hence, for $\hat{\tau}_{S_K}^2 > 0$, we have

$$(|A_0| - \sum_{i\in A_1} \lambda((\frac{l_i}{N}\alpha_1) \wedge \alpha_2))\hat{\tau}_{S_K}^2 \leq |A_0|\tau_*^2 + \sum_{i\in A_1} \lambda((\frac{l_i}{N}\alpha_1) \wedge \alpha_2)\hat{\sigma}_i^2 \tag{2}$$

We know that $\hat{\sigma}_i^2$ is $\frac{\chi_m^2}{m}\sigma_i^2$-distributed with mean at most C and the variance at most $\frac{C}{m}$. Apply the the Chebyshev's inequality to the quantity $\sum_{i \in i\in A} a_i\hat{\sigma}_i^2$ for set A and coefficient sequence $\{a_i\}$ we have

$$P(\sum_{i\in A_1} a_i\hat{\sigma}_i^2 \leq C(\sum_{i\in A_1} a_i + \sqrt{\frac{\sum_{i\in A_1} a_i^2 \log N}{m}})) \to 1 \tag{3}$$

Based on Lemma A.1, there is a constant c large enough such that for all $l = 1, 2, \ldots, |A_1|$, we have $\lambda(\frac{l\alpha_1}{N}) \leq c(\frac{N}{\alpha_1 l})^{\frac{2}{5}}$, and

$$\sum_{l=1}^{|A_1|} \lambda(\frac{l\alpha_1}{N}) \leq c\sum_{l=1}^{|A_1|} (\frac{N}{l\alpha_1})^{\frac{2}{5}} \leq c(\frac{N}{\alpha_1})^{\frac{2}{5}} \int_0^{N\gamma} l^{-\frac{2}{5}} dl = \frac{5c}{3\alpha_1^{\frac{2}{5}}} N\gamma^{\frac{3}{5}}$$
$$\sum_{l=1}^{|A_1|} \lambda^2(\frac{l\alpha_1}{N}) \leq c^2\sum_{l=1}^{|A_1|} (\frac{N}{l\alpha_1})^{\frac{4}{5}} \leq c^2(\frac{N}{\alpha_1})^{\frac{4}{5}} \int_0^{N\gamma} l^{-\frac{4}{5}} dl = \frac{5c^2}{\alpha_1^{\frac{4}{5}}} N\gamma^{\frac{1}{5}}$$

Combine the above inequality with equation (3), we have

$$P\left(\sum_{i\in A_1} \lambda((\frac{l_i}{N}\alpha_1) \vee \alpha_2)\hat{\sigma}_i^2 < C\left(\frac{5c}{3\alpha_1^{\frac{2}{5}}} N\gamma^{\frac{3}{5}} + \gamma N\lambda(\alpha_2) + \sqrt{\frac{(5c^2\alpha_1^{-\frac{4}{5}} N\gamma^{\frac{1}{5}} + N\gamma\lambda^2(\alpha_2))}{m}\log N}\right)\right) \to 1$$

Let $c_1 := C\left(\frac{5c}{3\alpha_1^{\frac{2}{5}}} N\gamma^{\frac{3}{5}} + \gamma N\lambda(\alpha_2) + \sqrt{\frac{(5c^2\alpha_1^{-\frac{4}{5}} N\gamma^{\frac{1}{5}} + N\gamma\lambda^2(\alpha_2))}{m}\log N}\right)$. Recall that $M_0 = \{|\tau^2 - \tau_*^2| \leq (1 - \sqrt{\gamma})\Delta_1\}$ happens with probability approaching 1 from Lemma A.2. For any $\delta > 0$, we have

$$\lim_{N\to\infty} P(\hat{\tau}^2 \leq \tau^2 + \delta(\tau^2 + C), M_0)$$

$$\leq \lim_{N\to\infty} P\Big(\frac{(1-\gamma)(\tau^2+\Delta_1)+c_1}{1-\gamma\lambda(\alpha_2)-\frac{5a}{3}\gamma^{\frac{3}{5}}\alpha_1^{-\frac{2}{5}}} < \tau^2+\delta(\tau^2+C)\Big) - \lim_{N\to\infty} P(M_0^c) - \lim_{N\to\infty} P\Big(\sum_{i\in A_1}\lambda\Big(\Big(\frac{l_i}{N}\alpha_1\Big)\vee\alpha_2\Big)\hat{\sigma}_i^2 > c_1\Big)$$

$$= \lim_{N\to\infty} P\Big(\frac{(1-\gamma)(\tau^2+\Delta_1)+c_1}{1-\gamma\lambda(\alpha_2)-\frac{5a}{3}\gamma^{\frac{3}{5}}\alpha_1^{-\frac{2}{5}}} < \tau^2+\delta(\tau^2+C)\Big)$$

$$= \lim_{N\to\infty} P\Big(\frac{(\gamma\lambda(\alpha_2)+\frac{5a}{3}\gamma^{\frac{3}{5}}\alpha_1^{-\frac{2}{5}}-\gamma)\tau^2+(1-\gamma)\Delta_1+c_1}{1-\gamma\lambda(\alpha_2)-\frac{5a}{3}\gamma^{\frac{3}{5}}\alpha_1^{-\frac{2}{5}}} < \delta(\tau^2+C)\Big) = 1$$

□

Proof of Theorem 2.8

Proof. Let R_0 and R_1 be the number of rejected nulls and non-nulls using level α_1(note that R_1 is the R_K^1 in Theorem 2.5). Define $V_i = \mathbb{1}_{\{H_i\ rejected\}}$ for each $i \in A_0$, p_i be the p values calculated using $\hat{\tau}^2$. We can express the FDR as

$$FDR = \sum_{l=0}^{|A_1|}\sum_{l_0=1}^{|A_0|} E\Big[\mathbb{1}_{R_1=l}\mathbb{1}_{R_0=l_0}\frac{\sum_{i\in A_0} V_i}{l+l_0}\Big] = \sum_{l=0}^{|A_1|}\sum_{l_0=1}^{|A_0|} E\Big[\mathbb{1}_{R_1=l}\mathbb{1}_{R_0=l_0}\frac{\sum_{i\in A_0}\mathbb{1}_{p_i\leq(\frac{l+l_0}{N}\alpha_1)\wedge\alpha_2}}{l+l_0}\Big]$$

We can further decompose the expression for FDR into two parts

$$\begin{aligned} FDR = &\underbrace{\sum_{l=0}^{|A_1|}\sum_{l_0=1}^{|A_0|} E\Big[\mathbb{1}_{R_1=l}\mathbb{1}_{R_0=l_0}\mathbb{1}_{\hat{\tau}^2\geq\tau_l^2}\frac{\sum_{i\in A_0}\mathbb{1}_{p_i\leq(\frac{l+l_0}{N}\alpha_1)\wedge\alpha_2}}{l+l_0}\Big]}_{I_1} \\ &+\underbrace{\sum_{l=0}^{|A_1|}\sum_{l_0=1}^{|A_0|} E\Big[\mathbb{1}_{R_1=l}\mathbb{1}_{R_0=l_0}\mathbb{1}_{\hat{\tau}^2<\tau_l^2}\frac{\sum_{i\in A_0}\mathbb{1}_{p_i\leq(\frac{l+l_0}{N}\alpha_1)\wedge\alpha_2}}{l+l_0}\Big]}_{I_2} \end{aligned}$$

By Theorem 2.5, we know $I_2 \leq P(\cup_{l=0}^{|A_1|}\{R_1 = l, \hat{\tau}^2 < \tau_l^2\}) \to 0$, and we need only to bound I_1. Let $\mathcal{F}_i = \{x_1^2, \ldots, x_{i-1}^2, x_{i+1}^2, \ldots, x_N^2, \hat{\sigma}_1^2, \ldots, \hat{\sigma}_{i-1}^2, \hat{\sigma}_{i+1}^2, \ldots, \hat{\sigma}_N^2\}$. Notice that

- Let us take x_i^2 and $\hat{\sigma}_i^2$ and set their value to ∞ and 0, and denote new number of rejections for the null and non-null by $\tilde{R}_0$ and $\tilde{R}_1$. If p_i is rejected, we know $\hat{\tau}^2$ is not calculated using x_i^2 or $\hat{\sigma}_i^2$. This new number of rejections is exactly R_0 and R_1 if we have rejected hypothesis i:

$$\{R_1 = l, R_0 = l_0, p_i \leq \frac{\alpha_1(l+l_0)}{N}\wedge\alpha_2, \hat{\tau}^2 \geq \tau_l^2\} = \{\tilde{R}_1 = l, \tilde{R}_0 = l_0, p_i \leq \frac{\alpha_1(l+l_0)}{N}\wedge\alpha_2, \hat{\tau}^2 \geq \tau_l^2\}$$

We take the expectation conditional on $\mathcal{F}_i$:

$$\begin{aligned}
I_1 &= \sum_{i\in A_0}\sum_{l=0}^{|A_1|}\sum_{l_0=1}^{|A_0|} E[\mathbb{1}_{\tilde{R}_1=l}\mathbb{1}_{\tilde{R}_0=l_0}E[\frac{\mathbb{1}_{p_i\leq\frac{\alpha_1(l+l_0)}{N}\wedge\alpha_2}}{(l+l_0)}\mathbb{1}_{\hat{\tau}^2\geq\tau_l^2}|\mathcal{F}_i]] \\
&\leq \sum_{i\in A_0}\sum_{l=0}^{|A_1|}\sum_{l_0=1}^{|A_0|} E[\mathbb{1}_{\tilde{R}_1=l}\mathbb{1}_{\tilde{R}_0=l_0}E[\frac{\mathbb{1}_{p_{i,l}\leq\frac{\alpha_1(l+l_0)}{N}\wedge\alpha_2}}{(l+l_0)}|\mathcal{F}_i]] \\
&= \sum_{i\in A_0}\sum_{l=0}^{|A_1|}\sum_{l_0=1}^{|A_0|} E[\mathbb{1}_{\tilde{R}_1=l}\mathbb{1}_{\tilde{R}_0=l_0}E[\frac{\mathbb{1}_{p_{i,l}\leq\frac{\alpha_1(l+l_0)}{N}\wedge\alpha_2}}{(l+l_0)}]] \qquad (4)
\end{aligned}$$

By Lemma A.1, part (2), we know $\lim_{N\to\infty}\sup_{i\in A_0}\sup_{l\geq 0}\sup_{\alpha\geq\min\left(\alpha_2,\frac{(l\vee 1)\alpha_1}{N}\right)}\frac{P(p_{i,l}\leq\alpha)}{\alpha}=1$. As a result, for any $\delta>0$, there exists a N_0 such that for all $N>N_0$, we have

$$\sup_{l\geq 0}\sup_{l_0\geq 1} P(p_{i,l}\leq(\frac{l+l_0}{N}\alpha_1)\wedge\alpha_2)\leq(1+\delta)(\frac{l+l_0}{N}\alpha_1)\wedge\alpha_2$$

Rearrange the righthand side of equation (4), we have $I_1\leq(1+\delta)\alpha_1$ for any $\delta>0$. Hence $\lim_{N\to\infty} I_{1,1}\leq\alpha_1$ and $\lim_{N\to\infty} FDR\leq\alpha_1$. □

Proof of Theorem 2.10

Proof. From Theorem 2.7, for any $\delta_1>0$, we know $M=\{\hat{\tau}^2\leq\tau^2+\delta_1(\tau^2+C)\}$ happens with probability approaching 1, which leads to the following result:

$$\begin{aligned}
P(\phi_{i,\alpha}=0) \leq& P(x_i^2\leq\widetilde{F}_i^{-1}(\alpha)(\hat{\tau}^2+\hat{\sigma}_i^2),M)+P(M^c) \\
\leq& \underbrace{P(z_i^2\leq\frac{\widetilde{F}_i^{-1}(\alpha)(\tau^2+\hat{\sigma}_i^2+\delta_1(\tau^2+C))}{\tau_i^2+\sigma_i^2})}_{I_{i,\alpha}}+P(M^c)
\end{aligned}$$

We now prove that $I_{i,\alpha}$ is no much larger than the oracle loss. We know that there exists a constant $f_{max}\geq\sup_t\frac{dP(z_i^2\leq t)}{dt}$, and for any δ_2, there is a constant w large enough such that

$$\sup_{\delta>0}\sup_{i\in A_1}\frac{P(z_i^2\leq w(1+\delta))}{P(z_i^2\leq w)(1+\delta)}\leq 1+\delta_2$$

For any α, we either have $\frac{\tilde{F}_i^{-1}(\alpha)(\tau^2+C)}{\tau_i^2+\sigma_i^2} \leq \frac{w}{\sqrt{\delta_1}}$ or not. If $\frac{\tilde{F}_i^{-1}(\alpha)(\tau^2+C)}{\tau_i^2+\sigma_i^2} \leq \frac{w}{\sqrt{\delta_1}}$, we have $I_{i,\alpha} \leq P(\phi_i^* = 0) + wf_{max}\sqrt{\delta_1}$. If$\frac{\tilde{F}_i^{-1}(\alpha)(\tau^2+C)}{\tau_i^2+\sigma_i^2} > \frac{w}{\sqrt{\delta_1}}$, we have

$$\begin{aligned}
I_{i,\alpha} \leq &\int_{\sqrt{\delta_1}C}^{\infty} P(z_i^2 \leq \frac{\widetilde{F}_i^{-1}(\alpha)(\tau^2+y)}{\tau_i^2+\sigma_i^2})(\frac{\tau^2+y+\delta_1(\tau^2+C)}{\tau^2+y})f_{\hat{\sigma}_i^2}(y)dy + P(\hat{\sigma}_i^2 \leq \sqrt{\delta_1}C) + \delta_2 \\
\leq &P(\phi_{i,\alpha}^* = 0) + \underbrace{\int_0^{\infty} P(\bar{x}_i^2 \leq \widetilde{F}_i^{-1}(\alpha)(\tau^2+y))\frac{\delta_1(C+\tau^2)}{y+\tau^2}f_{\hat{\sigma}_i^2}(y)dy}_{I_1} + P(\hat{\sigma}_i^2 \leq \delta_1\sigma_i^2) + \delta_2
\end{aligned}$$

In the integral I_1, because $P(x_i^2 \leq \widetilde{F}_i^{-1}(\alpha)(\tau^2+\hat{\sigma}_i^2+y))$ is an increasing function in y while $\frac{\delta_1(C+\tau^2)}{y+\tau^2}$ is a decreasing function in y, we have

$$I_1 \leq P(\phi_i^* = 0)\int_0^{\infty}\frac{\delta_1(C+\tau^2)}{y+\tau^2}f_{\hat{\sigma}_i^2}(y)dy$$

$\hat{\sigma}_i^2$ is $\frac{\chi_m^2}{m\sigma_i^2}$ distributed, the expectation of its inverse is $\frac{m}{(m-2)}$, Recall that $\min\sigma_i^2 = 1$:

$$\int_0^{\infty}\frac{\delta_1(C+\tau^2)}{y+\tau^2}f_{\hat{\sigma}_i^2}(y)dy \leq \int_0^{\infty}(\frac{\delta_1 C}{y}+\delta_1)f_{\hat{\sigma}_i^2}(y)dy = \delta_1(1+\frac{Cm}{m-2})$$

As a result, we have

$$P(\phi_{i,\alpha} = 0) - P(\phi_i^* = 0) \leq \max_{i\in A_1}(\sqrt{\delta_1}wf_{max}, \delta_1 + \delta_1\frac{Cm}{m-2} + P(\hat{\sigma}_i^2 \leq \sqrt{\delta_1}C) + \delta_2) + P(M^c)$$

The right-hand-side of the above expression does not depend of i or α. For any $\delta > 0$, we can take N large enough and δ_1, δ_2 small enough such that

$$\delta_1 + \delta_1\frac{Cm}{m-2} + \max_{i\in A_1}P(\hat{\sigma}_i^2 \leq \sqrt{\delta_1}C) + \delta_2 + P(M^c) < \delta, \quad \sqrt{\delta_1}wf_{max} + P(M^c) < \delta$$

Hence, we have

$$\lim_{N\to\infty}\sup_{i\in A_1}\sup_{\alpha\geq 0}(P(\phi_{i,\alpha} = 0) - P(\phi_{i,\alpha}^* = 0)) \leq 0$$

or

$$\lim_{N\to\infty}\inf_{i\in A_1}\inf_{\alpha\geq 0}(P(\phi_{i,\alpha} = 1) - P(\phi_{i,\alpha}^* = 1)) \geq 0$$

□

B Proof of Lemmas A.1, A.2

Proof of Lemma A.1

Proof. (1) Let $1-\tilde{T}_m(.)$ be the cumulative function of a t distribution with m degree of freedom and $1-\tilde{\Phi}(.)$ be the cumulative function of a normal. Let $t_m(.)$ and $\phi(.)$ be there density function. We first show that for any fixed value $t \geq 0$, we have

$$2\tilde{\Phi}(\sqrt{t}) \leq \tilde{F}_i(t) \leq 2\tilde{T}_m(\sqrt{t}) \tag{5}$$

Let $a(\tau^2) := P(\frac{x_i^2}{\tau^2+\hat{\sigma}_i^2} \geq t) = P(\frac{(\tau^2+\sigma_i^2)z^2}{\tau^2+\sigma_i^2 u} \geq t)$, where z be a random variable with standard normal distribution and u be a random variable distributed as $\frac{\chi_m^2}{m}$, z and u are independent. The function $a(\tau^2)$ has a non-positive first derivative with respect to τ^2:

$$\begin{aligned}
\frac{da(\tau^2)}{d\tau^2} &= \frac{d}{d\tau^2}\int_0^\infty \int_{z^2 \geq \frac{t(\tau^2+u\sigma_i^2)}{\tau^2+\sigma_i^2}} \phi(z)dz f_u(u)du \\
&= \int_{u=0}^\infty f_u(u)\phi(\sqrt{\frac{t(\tau^2+u\sigma_i^2)}{\tau^2+\sigma_i^2}})\sqrt{\frac{\tau^2+\sigma_i^2}{t(\tau^2+u\sigma_i^2)}}\frac{\sigma_i^2 t(u-1)}{(\tau^2+\sigma_i^2)^2}du \\
&\propto \int_{u=0}^\infty f_u(u)e^{-\frac{t(\tau^2+u\sigma_i^2)}{\tau^2+\sigma_i^2}}\sqrt{\frac{1}{\tau^2+u\sigma_i^2}}(u-1)du
\end{aligned}$$

The expected value of u is 1: $\int_{x=0}^\infty f_u(x)(u-1) = 0$ and $e^{-\frac{t(\tau^2+u\sigma_i^2)}{\tau^2+\sigma_i^2}}\sqrt{\frac{1}{\tau^2+u\sigma_i^2}}$ is a decreasing function of u, thus $a(\tau^2)$ has a non-positive first derivative with respect to τ^2. For any fixed $t \geq 0$, we have

$$a(\infty) \leq P(\frac{x_i^2}{\tau^2+\hat{\sigma}_i^2} \geq t) \leq a(0)$$

We use the fact that $a(\infty) = 2\tilde{\Phi}(\sqrt{t})$ and $a(0) = 2\tilde{T}_m(\sqrt{t})$ to get equation (5). It is also easy to check that for any fixed non-negative t, $\tilde{T}_m(t)$ is non-increasing in m because when $m_1 < m_2$, the density ratio between the t-distribution with degree of freedom m_1 and that with degree of freedom m_2 is non-decreasing in the positive part and non-increasing in the negative part. As a result, $\tilde{T}_m(t)$ is non-increasing in m for any fixed t and $\tilde{F}_i(t) \leq 2\tilde{T}_5(\sqrt{t})$. Apply the Mill's ratio result for the t-distribution(Soms (1976)):

$$\tilde{T}_m(t) < \frac{t_m(t)}{t}(1+\frac{t^2}{m}) \tag{6}$$

we have

$$\tilde{F}(t) \leq \frac{2\Gamma(\frac{m+1}{2})}{\sqrt{\pi m t}\Gamma(\frac{m}{2})}(1+\frac{t}{m})^{-\frac{m-1}{2}} < \sqrt{\frac{2}{\pi}}(\frac{t}{m})^{-\frac{m}{2}} \to \lambda(\frac{l\alpha_1}{N}) \lesssim (\frac{N}{l\alpha_1})^{\frac{2}{5}}$$

As a direct result, we have $\frac{\Delta_1}{\tau^2+1}\lambda(\frac{\alpha_1}{N}) \to 0$. Because $\Delta_{2,l} \leq \alpha_2 \log(\frac{1}{\alpha_2})$, we have $\lambda(\alpha_2)\alpha_2 \log\frac{1}{\alpha_2} \to 0$, hence $\sup_{l\geq 0} \Delta_{2,l}\lambda(\lambda_2) \to 0$. The result $\sup_{l\geq 1} \frac{\Delta_{2,l}}{\tau^2+1}\lambda(\frac{l\alpha_1}{N}) \to 0$ also holds because

- If $\frac{l}{N}$ is a positive constant, $\frac{\Delta_{2,l}}{\tau^2+1} \to 0$ because $\alpha_2 \to 0$.
- If $\frac{l}{N} \to 0$ and $\frac{l}{N} \gtrsim \frac{\log^2 N}{N}$, $\frac{\Delta_{2,l}}{\tau^2+1} \lesssim \frac{l}{N}\log\frac{N}{l}$, $\lambda(\frac{l\alpha_1}{N})\frac{\Delta_{2,l}}{\tau^2+1} \to 0$.
- If $\frac{l}{N} \lesssim \frac{\log^2 N}{N}$, $\frac{\Delta_{2,l}}{\tau^2+1} \lesssim \frac{\log^2 N}{N}\log\frac{N}{\log^2 N}$ and $N^{\frac{2}{5}}\frac{\Delta_{2,l}}{\tau^2+1} \to 0$, hence we have $\lambda(\frac{l\alpha_1}{N})\frac{\Delta_{2,l}}{\tau^2+1} \to 0$.

(2)Based on part (1) and the fact that $\Delta_{2,0} \leq \Delta_{2,1}$, let $\alpha_3 = \min\left(\alpha_2, \frac{(l\vee 1)\alpha_1}{N}\right)$, we have:

$$\sup_{l\geq 0}\sup_{\alpha\geq\alpha_3} \frac{\Delta_1+\Delta_{2,l}}{\tau^2+1}\lambda(\alpha) \to 0 \tag{7}$$

Because $\tau_l^2 \leq \tau^2$, we always have $\frac{P(p_{i,l}\leq\alpha)}{\alpha} \geq 1$, and we need only to check that, for any $\delta > 0$, $\frac{P(p_{i,l}\leq\alpha)}{\alpha} \leq 1+\delta$ holds uniformly for large N. We break the expression in the statement into two parts:

$$\sup_{i,l}\sup_{\alpha\geq\alpha_3} P(p_{i,l}\leq\alpha))/\alpha = \sup_{i,l,\alpha}\int_{y=0}^{\infty} P(x_i^2 \geq (y+\tau_l^2)\tilde{F}_i^{-1}(\alpha))f_{\hat{\sigma}_i^2}(y)dy/\alpha = \sup_{i,l,\alpha}(I_{1,i,l,\alpha}+I_{2,i,l,\alpha})$$

where $I_{1,i,l,\alpha} = \int_{\frac{y+\tau^2}{\tau^2+C}\tilde{F}_i^{-1}(\alpha)>\frac{1}{\delta}} P(x_i^2 \geq (y+\tau_l^2)\tilde{F}_i^{-1}(\alpha))f_{\hat{\sigma}_i^2}(y)dy/\alpha$ and $I_{2,i,l,\alpha} = \int_{\frac{y+\tau^2}{\tau^2+C}\tilde{F}_i^{-1}(\alpha)\leq\frac{1}{\delta}} P(x_i^2 \geq (y+\tau_l^2)\tilde{F}_i^{-1}(\alpha))f_{\hat{\sigma}_i^2}(y)dy/\alpha$, with δ being any positive constant. For $I_{1,i,l,\alpha}$, we use the following Mill's result for the normal(Gordon (1941))

$$\frac{t}{t^2+1}\phi(t) < \tilde{\Phi}(t) < \frac{\phi(t)}{t} \tag{8}$$

to upper bound $P(x_i^2 \geq (y+\tau_l^2)\tilde{F}_i^{-1}(\alpha))$ and lower bound $P(x_i^2 \geq (y+\tau^2)\tilde{F}_i^{-1}(\alpha))$ in terms of the density:

$$I_{1,i,l,\alpha} \leq \int_{\frac{y+\tau^2}{\tau^2+C}\tilde{F}_i^{-1}(\alpha)>\frac{1}{\delta}} P(x_i^2 \geq (y+\tau^2)\tilde{F}_i^{-1}(\alpha))\frac{\tau^2+\sigma_i^2+(y+\tau^2)\tilde{F}_i^{-1}(\alpha)}{(y+\tau_l^2)\tilde{F}_i^{-1}(\alpha)}e^{\frac{(\Delta_1+\Delta_{2,l})\tilde{F}_i^{-1}(\alpha)}{2(\tau^2+\sigma_i^2)}}f_{\hat{\sigma}_i^2}(y)dy/\alpha$$

$$\leq \int_{\frac{y+\tau^2}{\tau^2+C}\tilde{F}_i^{-1}(\alpha)>\frac{1}{\delta}} P(x_i^2 \geq (y+\tau^2)\tilde{F}_i^{-1}(\alpha))(1+\frac{\tau^2+\sigma_i^2+(\Delta_1+\Delta_{2,l})\tilde{F}_i^{-1}(\alpha)}{(y+\tau_l^2)\tilde{F}_i^{-1}(\alpha)})e^{\frac{(\Delta_1+\Delta_{2,l})\tilde{F}_i^{-1}(\alpha)}{2(\tau^2+\sigma_i^2)}} f_{\hat{\sigma}_i^2}(y)dy/\alpha$$

For $I_{2,i,l,\alpha}$:

$$\begin{aligned}
&I_{2,i,l,\alpha} - \int_{\frac{y+\tau^2}{\tau^2+C}\tilde{F}_i^{-1}(\alpha)\leq\frac{1}{\delta}} P(x_i^2 \geq (y+\tau^2)\tilde{F}_i^{-1}(\alpha)) f_{\hat{\sigma}_i^2}(y)dy/\alpha \\
=&\int_{\frac{y+\tau^2}{\tau^2+C}\tilde{F}_i^{-1}(\alpha)\leq\frac{1}{\delta}} \Big(P(x_i^2 \geq (y+\tau_l^2)\tilde{F}_i^{-1}(\alpha)) - P(x_i^2 \geq (y+\tau^2)\tilde{F}_i^{-1}(\alpha))\Big) f_{\hat{\sigma}_i^2}(y)dy/\alpha
\end{aligned}$$

Recall that $P(x_i^2 \geq (y+\tau^2)\tilde{F}_i^{-1}(\alpha)) = 2\tilde{\Phi}(\sqrt{\frac{(y+\tau^2)\tilde{F}_i^{-1}(\alpha)}{\tau^2+\sigma_i^2}})$ and $P(x_i^2 \geq (y+\tau_l^2)\tilde{F}_i^{-1}(\alpha)) = 2\tilde{\Phi}(\sqrt{\frac{(y+\tau_l^2)\tilde{F}_i^{-1}(\alpha)}{\tau^2+\sigma_i^2}})$, we can bound the difference by the product of the difference in the interval length and the upper bound of the normal density

$$2\tilde{\Phi}(\sqrt{\frac{(y+\tau_l^2)\tilde{F}_i^{-1}(\alpha)}{\tau^2+\sigma_i^2}})-2\tilde{\Phi}(\sqrt{\frac{(y+\tau^2)\tilde{F}_i^{-1}(\alpha)}{\tau^2+\sigma_i^2}}) \leq 2\sqrt{\frac{1}{2\pi}}(\sqrt{\frac{(y+\tau^2)\tilde{F}_i^{-1}(\alpha)}{\tau^2+\sigma_i^2}}-\sqrt{\frac{(y+\tau_l^2)\tilde{F}_i^{-1}(\alpha)}{\tau^2+\sigma_i^2}})$$

We know that for any positive value x, y, we have $\sqrt{x+y} \leq \sqrt{x}+\sqrt{y}$, as a result, we have

$$\frac{2\tilde{\Phi}(\sqrt{\frac{(y+\tau_l^2)\tilde{F}_i^{-1}(\alpha)}{\tau^2+\sigma_i^2}}) - 2\tilde{\Phi}(\sqrt{\frac{(y+\tau^2)\tilde{F}_i^{-1}(\alpha)}{\tau^2+\sigma_i^2}})}{2\tilde{\Phi}(\sqrt{\frac{(y+\tau^2)\tilde{F}_i^{-1}(\alpha)}{\tau^2+\sigma_i^2}})} \leq \frac{\sqrt{\frac{(\tau^2-\tau_l^2)\tilde{F}_i^{-1}(\alpha)}{2\pi(\tau^2+\sigma_i^2)}}}{\tilde{\Phi}(\sqrt{\frac{(y+\tau^2)\tilde{F}_i^{-1}(\alpha)}{\tau^2+\sigma_i^2}})}$$

In other words, we have

$$I_{2,i,l,\alpha} \leq \int_{\frac{y+\tau^2}{\tau^2+C}\tilde{F}_i^{-1}(\alpha)\leq\frac{1}{\delta}} P(x_i^2 \geq (y+\tau^2)\tilde{F}_i^{-1}(\alpha))(1+\frac{\sqrt{\frac{(\tau^2-\tau_l^2)\tilde{F}_i^{-1}(\alpha)}{2\pi(\tau^2+\sigma_i^2)}}}{\tilde{\Phi}(\sqrt{\frac{C}{\delta}})}) f_{\hat{\sigma}_i^2}(y)dy/\alpha$$

Combine them together and apply equation (7), we have

$$\sup_{i,l,\alpha}(I_{1,i,l,\alpha}+I_{2,i,l,\alpha}-1) \leq \sup_{i,l,\alpha}(\int_{\frac{y+\tau^2}{C+\tau^2}\tilde{F}^{-1}(\alpha)\leq\frac{1}{\delta}} P(x_i^2 \geq (y+\tau^2)\tilde{F}_i^{-1}(\alpha))\frac{\sqrt{\frac{(\Delta_1+\Delta_{2,l})\tilde{F}^{-1}(\alpha)}{2\pi(\tau^2+C)}}}{\tilde{\Phi}(\sqrt{\frac{C}{\delta}})} f_{\hat{\sigma}_i^2}(y)dy/\alpha$$

$$
\begin{aligned}
&+\int_{\frac{y+\tau^2}{C+\tau^2}\tilde{F}_i^{-1}(\alpha)>\frac{1}{\delta}} P(x_i^2 \geq (y+\tau^2)\tilde{F}_i^{-1}(\alpha))\frac{\tau^2+\sigma_i^2+(\Delta_1+\Delta_{2,l})\tilde{F}_i^{-1}(\alpha)}{(y+\tau_l^2)\tilde{F}_i^{-1}(\alpha)}e^{\frac{(\Delta_1+\Delta_{2,l})\tilde{F}_i^{-1}(\alpha)}{2(\tau^2+\sigma_i^2)}} f_{\hat{\sigma}_i^2}(y)dy/\alpha) \\
&\to \sup_{i,l,\alpha}\int_{\frac{y+\tau^2}{C+\tau^2}\tilde{F}_i^{-1}(\alpha)>\frac{1}{\delta}} P(x_i^2 \geq (y+\tau^2)\tilde{F}_i^{-1}(\alpha))\frac{\tau^2+\sigma_i^2}{(y+\tau_l^2)\tilde{F}_i^{-1}(\alpha)} f_{\hat{\sigma}_i^2}(y)dy/\alpha \\
&\leq \sup_{i,l,\alpha}\int_{\frac{y+\tau^2}{C+\tau^2}\tilde{F}_i^{-1}(\alpha)>\frac{1}{\delta}} P(x_i^2 \geq (y+\tau^2)\tilde{F}_i^{-1}(\alpha))\frac{\tau^2+C}{\frac{(\tau^2+C)}{\delta}-(\Delta_1+\Delta_{2,l})\tilde{F}_i^{-1}(\alpha)} f_{\hat{\sigma}_i^2}(y)dy/\alpha \\
&\to \sup_{i,\alpha}\int_{\frac{y+\tau^2}{C+\tau^2}\tilde{F}_i^{-1}(\alpha)>\frac{1}{\delta}} \delta P(x_i^2 \geq (y+\tau^2)\tilde{F}_i^{-1}(\alpha)) f_{\hat{\sigma}_i^2}(y)dy/\alpha < \delta
\end{aligned}
$$

As it holds for any $\delta > 0$, we have $\lim_{N\to\infty}\sup_{i\in A_0}\sup_{l\geq 0}\sup_{\alpha\geq\min\left((\alpha_2\vee\frac{\log^2 N}{N}),\frac{(l\vee 1)\alpha_1}{N}\right)}\frac{P(p_{i,l}\geq\alpha)}{\alpha} = 1$

(3) At $k+1^{th}$ iteration, for every point we removed, they need to satisfy that $x_i^2 \geq \tilde{F}_i^{-1}(\alpha_1)(\hat{\tau}_{S_k}^2+\hat{\sigma}_i^2)$. From equation (5), we have(recall that $\alpha_1 < \frac{1}{2e}$):

$$\tilde{F}^{-1}(\alpha_1) \geq (\tilde{\Phi}^{-1}(\frac{\alpha_1}{2}))^2 > 1$$

as a result, $x_i^2 - \hat{\sigma}_i^2 \geq \hat{\tau}_{S_k}^2 \Rightarrow$ the τ^2 estimate is non-increasing. □

Proof of Lemma A.2

Proof of M_0 happening with probability approaching one: We know that $\hat{\sigma}_i^2 \sim \sigma_i^2\frac{\chi_m^2}{m}$ and $x_i^2 \sim (\tau^2+\sigma_i^2)\chi_1^2$. Because $\sigma_i^2 \leq C$, we have $E[x_i^2-\hat{\sigma}_i^2] = \tau^2$ and $Var[x_i^2-\hat{\sigma}_i^2] \leq \tau^2+(1+\frac{1}{m})C$. Result follows from the Chebyshev's inequality.
Proof of M_1 happening with probability approaching one: For the event M_1, consider the event $A_{k,l} := \{|B_{1,l}| = k\}$. Use Lemma A.1 part (2), and take $\delta < \frac{1}{2e\alpha_1}-1$, for large enough N:

$$\sup_{i\in A_0}\sup_{l\geq 0,k\geq 1}\frac{P(p_{i,l}\leq\frac{(l+k)\alpha_1}{N})}{\frac{(l+k)\alpha_1}{N}} < 1+\delta \tag{9}$$

Event $A_{k,l}$ is contained in the event that there are k null p-values at most $\frac{(l+k)\alpha_1}{N}$, hence, $P(A_{k,l}) \leq \binom{|A_0|}{k}(\frac{l+k}{N}(1+\delta)\alpha_1)^k$. Let $k_l := \lceil |A_0|\max(\frac{l}{N},\frac{\log^2 N}{N})\rceil$, for $l = 0,1,\ldots,|A_1|$, we have

$$P(M_1^c) \leq \sum_{l=0}^{|A_1|} \sum_{k \geq k_l} \left(\begin{array}{c} |A_0| \\ k \end{array} \right) (\frac{l+k}{N}(1+\delta)\alpha_1)^k$$

Let $a_{k,l} = \left(\begin{array}{c} |A_0| \\ k \end{array} \right) (\frac{k+l}{N}(1+\delta)\alpha_1)^k$, $u_{k,l} = \frac{k+l}{k}$. It is easy to check that $u_{k,l}$ is decreasing in k and $x \mapsto xe^{1/x}$ is increasing on $[1, \infty]$. For $k \geq k_l \geq (1-\gamma)l$, we have that $u_{k,l} \leq \frac{2}{1-\gamma}$, and $(1+\frac{1}{k+l})^k \leq e^{1/u_{k,l}}$, as $\log(1+\frac{1}{k+l})^k = k\log(1+\frac{1}{k+l}) \leq \frac{k}{k+l}$. Hence, for large N and any l considered, $a_{k,l}$ is non-increasing in k when $k \geq k_l$:

$$\begin{aligned}
\sup_{l,k} \frac{a_{k+1,l}}{a_{k,l}} =& \sup_{l,k} \frac{(|A_0|-k)}{k+1} \frac{(k+l+1)(1+\delta)\alpha_1}{N} (1+\frac{1}{k+l})^k \\
\leq& \sup_{l,k} (1-\gamma)(1+\delta)\alpha_1 u_{k,l} e^{\frac{1}{u_{k,l}}} \\
\leq& (1-\gamma)(1+\delta)\alpha_1 \frac{2}{1-\gamma} e^{\frac{1-\gamma}{2}} < 2\sqrt{e}(1+\delta)\alpha_1 < 1
\end{aligned}$$

Using sterling's approximations to upper bound a_{l,k_l}, the probability of M_1^c can be bounded as

$$\begin{aligned}
P(M_1^c) \leq \sum_{l=0}^{|A_1|} |A_0| a_{l,\lceil k_l \rceil} &\leq \sum_{l=0}^{|A_1|} |A_0| \frac{|A_0|^{k_l+|A_0|-k_l}}{k_l^{k_l}(|A_0|-k_l)^{|A_0|-k_l}} (\frac{k_l+l}{N}(1+\delta)\alpha_1)^{k_l} \\
&= \sum_{l=0}^{|A_1|} |A_0| \frac{|A_0|^{|A_0|-k_l}}{(|A_0|-k_l)^{|A_0|-k_l}} (\frac{|A_0|}{N})^{k_l} (\frac{k_l+l}{k_l}(1+\delta)\alpha_1)^{k_l} \\
&= \sum_{l=0}^{|A_1|} |A_0| \exp(k_l \log(u_{k_l,l}(1-\gamma)(1+\delta)\alpha_1) + (|A_0|-k_l)\log\frac{|A_0|}{|A_0|-k_l}) \\
&\leq \sum_{l=0}^{|A_1|} |A_0| \exp\left(k_l \log(2(1+\delta)\alpha_1) + (|A_0|-k_l)\log\frac{|A_0|}{|A_0|-k_l}\right)
\end{aligned}$$

The quantity inside the exponential is a decreasing function of k_l, as its derivative is

$$(\log(2(1+\delta)\alpha_1) + 1 + \log(\frac{|A_0|-k_l}{|A_0|}) < 0$$

Thus, for all l, it is less than or equal to its value at $k_l = \lceil \frac{\log^2 N}{N}|A_0|\rceil$. For $k_l = \lceil \frac{\log^2 N}{N}|A_0|\rceil$, we have $(|A_0|-k_l)\log\frac{|A_0|}{|A_0|-k_l} = k_l(o(1)+1)$, and $P(M_1) \geq 1-N^2 e^{k_l(\log(2(1+\delta)\alpha_1)+1+o(1))} \to 1$.
Proof of M_2 happening with probability approaching one: For the event M_2, we only need to check $B_{2,|A_1|}$ because $B_{2,l} \subseteq B_{2,l'}$ for all $l \leq l'$. By Lemma A.1 part (2), $\sup_i \sup_l \frac{P(p_{i,l}\leq\alpha_2)}{\alpha_2} \leq (1+\delta)$. As a result, $|B_{2,|A_1|}|$ is at most y ∼Bin($|A_0|$, $(1+\delta)\alpha_2$). The variable y has mean $(1+\delta)\alpha_2|A_0|$ and variance bounded by $(1+\delta)\alpha_2|A_0|$. We apply Chebyshev's inequality and reach our conclusion $P(M_2) = P(|B_{1,N\gamma}| \leq 2\alpha_2|A_0|) \geq P(y \leq 2\alpha_2|A_0|) \geq 1-\frac{(1+\delta)}{(1-\delta)^2\alpha_2|A_0|} \to 1$.
Proof of M_3 happening with probability approaching one: Let $\tilde{x}^2_\alpha$ be the upper α^{th} quantile of $\{x_i^2,\ i\in A_0\}$. It is sufficient to consider A_α, the set of x_i^2 whose value is no smaller than $\tilde{x}_\alpha$, so $|A_\alpha| = \lceil\alpha|A_0|\rceil$. Let $D_1 = \{\forall l = 0,1,\ldots,|A_1|, \frac{\tilde{x}^2_{t_l}}{2(\tau^2+C)\log\frac{1}{t_l}} \leq 1\}$. Let z be a standard normal variable. For each $i\in A_0$, we have $P(x_i^2 \geq 2(\tau^2+C)\log\frac{1}{t_l}) \leq P(z^2 \geq 2(\tau^2+C)\log\frac{1}{t_l}) = 2\tilde{\Phi}(\sqrt{2\log\frac{1}{t_l}}) \overset{eq.(8)}{\leq} \frac{t_l}{\sqrt{\pi\log\frac{1}{t_l}}}$. Because $|A_0| > cN$ for some positive constant c and $t_l \in [\frac{\log^2 N}{N}, \alpha_2)$, we apply the sterling's approximations:

$$P(D_1^c) \leq N\max_{l=0}^{|A_1|}\begin{pmatrix}|A_0|\\ \lceil|A_0|t_l\rceil\end{pmatrix}P(x_i^2 \geq 2(\tau^2+C)\log\frac{1}{t_l})^{\lceil|A_0|t_l\rceil} \leq \max_{l=0}^{|A_1|}\frac{N(\frac{t_l}{\sqrt{\pi\log\frac{1}{t_l}}})^{|A_0|t_l}}{t_l^{|A_0|t_l}(1-t_l)^{|A_0|-|A_0|t_l}}$$

As t_l goes to 0 in l, for large enough N, we have

$$\frac{1}{t_l^{|A_0|t_l}(1-t_l)^{|A_0|(1-t_l)}} = \exp\left(-|A_0|t_l\log t_l - |A_0|(1-t_l)\log(1-t_l)\right) \leq \exp(-|A_0|t_l\log t_l + 2|A_0|t_l)$$

As $t_l \geq \frac{\log^2 N}{N}$, $|A_0| > cN$, for N large enough, we have $P(D_1^c) \leq N\exp(|A_0|t_l(2-\frac{1}{2}\log\log\frac{1}{t_l})) \to 0$.

Now we show M_3 happens high probability. As $t_l \geq \frac{\log^2 N}{N}$ and $|A_0| > cN$, for N large enough, we have $0.3(\tau^2+C)t_l|A_0|\log\frac{1}{t_l} \geq 8(\tau^2+C)\log N$. Let $M_3' = \{\forall l = 0,1,\ldots,|A_1|,\ \sum_{i\in A_{t_l}} x_i^2 \leq 2.2(\tau^2+C)t_l|A_0|\log\frac{1}{t_l} + 8(\tau^2+C)\log N\}$, we have $M_3' \subseteq M_3$ for large N. Let $\tilde{A}_{t_l}$ be nulls such that $x_i^2 \geq 2(\tau^2+C)\log\frac{1}{t_l}$. When D_1 is true, if x_i^2 exceeds $2(\tau^2+C)\log\frac{1}{t_l}$, it must also exceeds $\tilde{x}^2_{t_l}$, in other words, $\tilde{A}_{t_l} \subseteq A_{t_l}$. Since $|A_{t_l}| = \lceil t_l|A_0|\rceil$ and $x_i^2 - 2.2(\tau^2+C)\log\frac{1}{t_l} < 0$ for $i \neq \tilde{A}_{t_l}$, for N large enough:

$$\{M_2',\ D_1\} = \{\forall l = 0,1,\ldots,|A_1|,\ \sum_{i\in A_{t_l}} x_i^2 \leq 2.2(\tau^2+C)t_l|A_0|\log\frac{1}{t_l} + 8(\tau^2+C)\log N, D_1\}$$

$$\subseteq\{\forall l=0,1,\ldots,|A_1|, \underbrace{\sum_{i\in\tilde{A}_{t_l}}(x_i^2-2.2(\tau^2+C)\log\frac{1}{t_l}))}_{I_l}\leq 8(\tau^2+C)\log N\}\cap D_1$$

Let $w_i=\frac{\tau^2+C}{\tau^2+\sigma_i^2}\in[1,C]$, $z_i=\frac{x_i}{\sqrt{\tau^2+\sigma_i^2}}\sim N(0,1)$ for $i\in A_0$. Rearrange I_l:

$$I_l=\sum_{i\in A_0}(x_i^2-2.2(\tau^2+C)\log\frac{1}{t_l})\mathbb{1}_{x_i^2\geq 2(\tau^2+C)\log\frac{1}{t_l}}=(\tau^2+C)\sum_{i\in A_0}\frac{1}{w_i}(z_i^2-2.2w_i\log\frac{1}{t_l})\mathbb{1}_{z_i^2\geq 2w_i\log\frac{1}{t_l}}$$

Let $y_i:=\frac{1}{w_i}(z_i^2-2.2w_i\log\frac{1}{t_l})\mathbb{1}_{z_i^2\geq 2w_i\log\frac{1}{t_l}}$. The moment generating function of y_i is($\lambda<\frac{w_i}{2}$):

$$\begin{aligned}M_{y_i}(\lambda)=&2\int_{z_i\geq\sqrt{2w_i\log\frac{1}{t_l}}}\exp(\frac{1}{w_i}(z_i^2-2.2w_i\log\frac{1}{t_l})\lambda)\frac{1}{\sqrt{2\pi}}\exp(-\frac{z_i^2}{2})dz_i+P(z_i^2\leq 2w_i\log\frac{1}{t_l})\\=&\frac{2}{\sqrt{1-2\frac{\lambda}{w_i}}}\exp(-2.2\lambda\log\frac{1}{t_l})\tilde{\Phi}(\sqrt{(1-2\frac{\lambda}{w_i})2w_i\log\frac{1}{t_l}})+(1-2\tilde{\Phi}(\sqrt{2w_i\log\frac{1}{t_l}}))\end{aligned}$$

By the Mill's ratio bound (8), we have

$$M_{y_i}(\lambda)\leq 1+2t_l^{w_i}(\frac{t_l^{0.2\lambda}}{(1-2\frac{\lambda}{w_i})\sqrt{4w_i\pi\log\frac{1}{t_l}}}-\frac{\sqrt{2w_i\log\frac{1}{t_l}}}{\sqrt{2\pi}(1+2w_i\log\frac{1}{t_l})})$$

Take $\lambda=\frac{1}{4}$. Because we have $t_l\to 0$ over l, for N large enough, we have $M_{y_i}(\frac{1}{4})\leq 1$. As a result, for N large enough: $P(I_l\geq 8(\tau^2+C)\log N)=P(\sum_{i\in A_0}y_i\geq 8\log N)\leq\frac{\prod_{i\in A_0}(M_{y_i}(\frac{1}{4}))}{\exp(2\log N)}\leq\frac{1}{N^2}$, and

$$P(\max_{l=0}^{|A_1|}I_l\leq 8(\tau^2+C)\log N)\to 1\Rightarrow P(M_3')\to 1\Rightarrow P(M_3)\to 1$$

Proof of M_4 and M_5 happening with probability approaching one: For a small constant c, we define $I_{1,l}=\sum_{i\in A_1}(x_i^2-\hat{\sigma}_i^2-\tau_l^2)\mathbb{1}_{\hat{\sigma}_i^2\geq c}\mathbb{1}_{x_i^2\leq\tilde{F}_i^{-1}(\alpha_2)(\tau_l^2+\hat{\sigma}_i^2)}$ and $I_{2,l}=\sum_{i\in A_1}(x_i^2-\hat{\sigma}_i^2-\tau_l^2)\mathbb{1}_{\hat{\sigma}_i^2<c}\mathbb{1}_{x_i^2\leq\tilde{F}_i^{-1}(\alpha_2)(\tau_l^2+\hat{\sigma}_i^2)}$. We want to show $I_{1,l}+I_{2,l}\geq-(\tau^2+1)(1-\gamma)\sqrt{|A_1|\log N}$ for all l with high probablity. Let ϵ and L be the constants in Assumption 2.2. We have

$\tau_l^2 \le \tau^2$ and for the smallest $\tau^2_{|A_1|}$, we have $\frac{\tau^2-\tau^2_{|A_1|}}{\tau^2+1} \to 0$. As $\alpha_2 \to 0$, for large enough N, we have $(\tilde{F}_i^{-1}(\alpha_2) - 1 - \epsilon)(\tau_l^2+1) > L(\tau^2+1)$ and

$$\{x_i^2 - (1+\epsilon)\hat{\sigma}_i^2 \le L(\tau^2+1)\} \cap \{\hat{\sigma}_i^2 \ge c\} \subseteq \{x_i^2 \le \tilde{F}_i^{-1}(\alpha_2)(\tau_l^2 + \hat{\sigma}_i^2)\} \cap \{\hat{\sigma}_i^2 \ge c\}$$

If we include any point in $\{x_i^2 - (1+\epsilon)\hat{\sigma}_i^2 \ge L(\tau^2+1)\}$, we increase $I_{1,l}$. Using also $\tau^2 \ge \tau_l^2$, we have $I_{l,1} \ge \sum_{i\in A_1}(x_i^2 - \hat{\sigma}_i^2 - \tau^2)\mathbb{1}_{\hat{\sigma}_i^2 \ge c}\mathbb{1}_{x_i^2-(1+\epsilon)\hat{\sigma}_i^2 \le L(\tau^2+1)}$. If we include any point in $\{x_i^2 - (\hat{\sigma}^2 + \tau_l^2) \ge 0\}$, we increase I_2. Thus, we have $I_2 \ge -(c+\tau_l^2)\sum_{i\in A_1}\mathbb{1}_{\hat{\sigma}_i^2<c}\mathbb{1}_{x_i^2-(\hat{\sigma}^2+\tau_l^2)\le 0} \ge -(c+\tau^2)\sum_{i\in A_1}\mathbb{1}_{\hat{\sigma}_i^2<c}\mathbb{1}_{x_i^2-(1+\epsilon)\hat{\sigma}^2\le L(\tau^2+1)}$. Hence, we have

$$I_{1,l} + I_{2,l} \ge \sum_{i\in A_1}[(x_i^2 - \hat{\sigma}_i^2 - \tau^2)\mathbb{1}_{\hat{\sigma}_i^2\ge c} - (\tau^2+c)\mathbb{1}_{\hat{\sigma}_i^2<c}]\mathbb{1}_{x_i^2-(1+\epsilon)\hat{\sigma}^2\le L(\tau^2+1)}$$

The lower bounds no longer involve l. Let $y_i = [(x_i^2 - \hat{\sigma}_i^2 - \tau^2)\mathbb{1}_{\hat{\sigma}_i^2\ge c} - (\tau^2+c)\mathbb{1}_{\hat{\sigma}_i^2<c}]\mathbb{1}_{x_i^2-(1+\epsilon)\hat{\sigma}_i^2\le L(\tau^2+1)}$ and let $\tilde{A} = \{i : x_i^2 - (1+\epsilon)\hat{\sigma}_i^2 \le L(\tau^2+1)\}$, we have $E[y_i] = P(\tilde{A})E[(x_i^2 - \hat{\sigma}_i^2 - \tau^2)\mathbb{1}_{\hat{\sigma}_i^2\ge c} - (\tau^2+c)\mathbb{1}_{\hat{\sigma}_i^2<c}|\tilde{A}]$. For any $y \ge 0$, we have

$$\begin{aligned}
E[\mathbb{1}_{\hat{\sigma}_i^2<y}|\tilde{A}] &= \int_{t\ge 0} P(\hat{\sigma}_i^2 \le y|(1+\epsilon)\hat{\sigma}_i^2 \ge t-(L+1))dP(x_i^2 > t) \\
&\le \int_{t\ge 0} P(\hat{\sigma}_i^2 \le y)dP(x_i^2>t) = P(\hat{\sigma}_i^2 \le y) \\
E[\hat{\sigma}_i^2|\tilde{A}] &= \int_{y\ge 0} y dP(\hat{\sigma}_i^2 > y|\tilde{A}) \ge \int_{y\ge 0} y dP(\hat{\sigma}_i^2 > y) = E[\hat{\sigma}_i^2]
\end{aligned}$$

$$E[x_i^2\mathbb{1}_{\hat{\sigma}_i^2<c}|\tilde{A}] = P(\hat{\sigma}_i^2<c)\int_{t=0}^{\infty} E[x_i^2|x_i^2 \le (1+\epsilon)t + L(\tau^2+1)]\frac{dP(\hat{\sigma}_i^2<t|\hat{\sigma}_i^2\le c)}{dP(\hat{\sigma}_i^2<t)}dP(\hat{\sigma}_i^2<t)$$

Because $E[x_i^2|x_i^2 \le (1+\epsilon)t + L(\tau^2+1)]$ is non-decreasing in t and $\frac{dP(\hat{\sigma}_i^2<t|\hat{\sigma}_i^2\le c)}{dP(\hat{\sigma}_i^2<t)}$ is non-increasing in t, we have

$$E[x_i^2\mathbb{1}_{\hat{\sigma}_i^2<c}|\tilde{A}] \le P(\hat{\sigma}_i^2<c)\int_{t=0}^{\infty} E[x_i^2|x_i^2 \le (1+\epsilon)t + L(\tau^2+1)]dP(\hat{\sigma}_i^2<t) = P(\hat{\sigma}_i^2<c)E[x_i^2|\tilde{A}]$$

We can now lower bound $E[(x_i^2 - \hat{\sigma}_i^2 - \tau^2)\mathbb{1}_{\hat{\sigma}_i^2\ge c}|\tilde{A}]$:

$$\begin{aligned}
E[(x_i^2 - \hat{\sigma}_i^2 - \tau^2)\mathbb{1}_{\hat{\sigma}_i^2\ge c}|\tilde{A}] &\ge E[x_i^2 - \hat{\sigma}_i^2|\tilde{A}] - E[x_i^2\mathbb{1}_{\hat{\sigma}_i^2<c}|\tilde{A}] - \tau^2 \\
&\ge E[x_i^2 - \hat{\sigma}_i^2|\tilde{A}] - P(\hat{\sigma}_i^2<c)E[x_i^2|\tilde{A}] - \tau^2
\end{aligned}$$

By Assumption 2.2, we have $E[x_i^2 - (1+\epsilon)\hat{\sigma}_i^2|\tilde{A}] \geq (1+\epsilon)\tau^2$, if we take c small enough such that $\max_{i\in A_1} P(\hat{\sigma}_i^2 < c) \leq \frac{\epsilon}{2(1+\epsilon)}$, we have

$$\begin{aligned}
E[y_i] &\geq E[x_i^2 - \hat{\sigma}_i^2|\tilde{A}] - P(\hat{\sigma}_i^2 < c)E[x_i^2|\tilde{A}] - \tau^2 \\
&\geq (1 - \frac{\epsilon}{2(1+\epsilon)})E[x_i^2 - (1+\epsilon)\hat{\sigma}_i^2|\tilde{A}] + \frac{\epsilon}{2}E[\hat{\sigma}_i^2|\tilde{A}] - \tau^2 \\
&\geq (1 - \frac{\epsilon}{2(1+\epsilon)})(1+\epsilon)\tau^2 - \tau^2 + \frac{\epsilon}{2} > 0
\end{aligned}$$

We can also bound $E[y_i^2]$:

$$E[y_i^2] = E[(x_i^2 - \hat{\sigma}_i^2 - \tau^2)^2 \mathbb{1}_{\hat{\sigma}_i^2 \geq c} \mathbb{1}_{\tilde{A}} + (\tau^2 + c)^2 \mathbb{1}_{\hat{\sigma}_i^2 < c} \mathbb{1}_{\tilde{A}}]$$

When $\tilde{A}$ is true, $(x_i^2 - \hat{\sigma}_i^2 - \tau^2)^2 \leq \max((\hat{\sigma}_i^2 + \tau^2)^2, (L(\tau^2+1) + \epsilon\hat{\sigma}_i^2)^2)$, therefore, we have

$$\begin{aligned}
E[y_i^2] &\leq (\tau^2 + c)^2 + E[(\hat{\sigma}_i^2 + \tau^2)^2] + E[(L(\tau^2+1) + \epsilon\hat{\sigma}_i^2)^2] \\
&\leq (\tau^2+1)^2 + (1+\epsilon^2)E[\hat{\sigma}_i^4] + (1+L^2)(\tau^2+1)^2 + 2(\tau^2 + L(\tau^2+1))E[\hat{\sigma}_i^2] \\
&\leq (1 + L^2 + (1+\epsilon^2)C^2(1 + \frac{1}{m}) + 2(L+2)C)(\tau^2+1)^2
\end{aligned}$$

We apply the Chebyshev's inequality to $\sum_{i\in A_1} y_i$:

$$P(\sum_{i\in A_1} y_i \geq -(1-\gamma)\sqrt{|A_1|\log N}(\tau^2+1)) \to 1$$

As we have $I_{1,l} + I_{2,l} \leq \sum_{i\in A_1} y_i$ holds for all l, hence, $P(M_4) \to 1$. For the event M_5, we have $\sum_{i\in A_1}(x_i^2 - \hat{\sigma}_i^2 - \tau^2) \leq \sum_{i\in A_1}(x_i^2 - \hat{\sigma}_i^2 - \tau^2)\mathbb{1}_{\tilde{A}} \leq \sum_{i\in A_1} y_i$. Therefore $P(M_5) \to 1$.

C Estimation procedures

In this section, we give the details of the truncated MLE estimate and the CM estimate of the spreading factor τ^2.

Truncated *MLE*: Let C be a normalization constant depending on the context, the likelihood function of the observed points from null distribution with mean level μ_i in terms of sufficient statistics $\bar{x}_i$ and $\hat{\sigma}_i^2(\hat{\sigma}_i^2 = m\hat{\sigma}_{\bar{x}_i}^2,\ \sigma_i^2 = m\sigma_{\bar{x}_i}^2)$ is

$$f = C\prod_{i\in A_0}(\tau^2)^{-\frac{1}{2}}\sigma_i^{-(m-k)}e^{-\frac{(\bar{x}_i-\mu_i)^2}{2\tau^2}}e^{-\frac{(m-k)\hat{\sigma}_i^2}{2\sigma_i^2}}$$

marginalized out μ_i:

$$f = C \prod_{i\in A_0} (\tau^2 + \sigma^2_{\bar{x}_i})^{-\frac{1}{2}} \sigma_i^{-(m-k)} \exp(-\frac{(m-k)\hat{\sigma}_i^2}{2\sigma_i^2} - \frac{\bar{x}_i^2}{2(\tau^2+\sigma^2_{\bar{x}_i})}) \tag{10}$$

For points in A_0 with mean difference $\bar{x}_i$ in $(-\delta_0, \delta_0)$, for a positive value δ_0, this truncated likelihood function is:

$$f_{truncated} = C \prod_{i\in A_0, \bar{x}_i \in(-\delta_0,\delta_0)} \frac{I_{[\bar{x}_i\in(-\delta_0,\delta_0)]}}{H(\tau^2, \sigma^2_{\bar{x}_i})} (\tau^2 + \sigma^2_{\bar{x}_i})^{-\frac{1}{2}} \sigma_i^{-(m-k)} \exp(-\frac{(m-k)\hat{\sigma}_i^2}{2\sigma_i^2} - \frac{\bar{x}_i^2}{2(\tau^2+\sigma^2_{\bar{x}_i})})$$

where

$$H(\tau, \sigma^2_{\bar{x}_i}) = \int_{x\in[-\delta_0,\delta_0]} \frac{1}{\sqrt{2\pi(\tau^2+\sigma^2_{\bar{x}_i})}} \exp(-\frac{x^2}{2(\tau^2+\sigma^2_{\bar{x}_i})})$$

Assuming that the observed $\{\bar{x}_i, \forall i \in A_1\}$ will not fall into the range $(-\delta_0, \delta_0)$, we have

$$f_{truncated} = C \prod_{\bar{x}_i \in(-\delta_0,\delta_0)} \frac{I_{[\bar{x}_i\in(-\delta_0,\delta_0)]}}{H(\tau^2, \sigma^2_{\bar{x}_i})} (\tau^2 + \sigma^2_{\bar{x}_i})^{-\frac{1}{2}} \sigma_i^{-(m-k)} \exp(-\frac{(m-k)\hat{\sigma}_i^2}{2\sigma_i^2} - \frac{\bar{x}_i^2}{2(\tau^2+\sigma^2_{\bar{x}_i})})$$

$$\begin{aligned} l_{truncated} &= -\log f_{truncated} \\ &= C + \sum_{i=1}^{N} I_{[\bar{x}_i\in(-\delta_0,\delta_0)]} (\log H_i + \frac{\log(\tau^2+\sigma^2_{\bar{x}_i})}{2} + \frac{m-k}{2}\log\sigma_i^2 + \frac{(m-k)\hat{\sigma}_i^2}{2\sigma_i^2} + \frac{\bar{x}_i^2}{2(\tau^2+\sigma^2_{\bar{x}_i})}) \end{aligned} \tag{11}$$

We can find the minimizer to the above target function by iteratively updating τ^2 and $\{\sigma_i^2 : -\delta_0 \le \bar{x}_i \le \delta_0\}$. We start from $\tau^2 = \tilde{\tau}^2 = 0$ and do the following,

For τ^2 fixed at $\tilde{\tau}^2$, find solutions to $\{\sigma_i^2\}$:

$$\tilde{\sigma}_i^2 = \arg\min_{\sigma_i^2} \log H_i + \frac{\log(\tau^2+\sigma^2_{\bar{x}_i})}{2} + \frac{m-k}{2}\log\sigma_i^2 + \frac{(m-k)\hat{\sigma}_i^2}{2\sigma_i^2} + \frac{\bar{x}_i^2}{2(\tau^2+\sigma^2_{\bar{x}_i})}$$

For $\{\sigma_i^2\}$ fixed at $\{\tilde{\sigma}_i^2\}$, find solution to τ^2 :

$$\tilde{\tau}_i^2 = \arg\min_{\tau^2} \sum_{i=1}^{N} I_{[\bar{x}_i\in(-\delta_0,\delta_0)]} (\log H_i + \frac{\log(\tau^2+\sigma^2_{\bar{x}_i})}{2} + \frac{\bar{x}_i^2}{2(\tau^2+\sigma^2_{\bar{x}_i})})$$

CM: The marginal density of $\bar{x}_i$(marginalized over the index i) for all genes can be written as following

$$f(x|\tau) \sim \frac{1}{N}\sum_{i\in A_0}\frac{e^{-\frac{x^2}{2(\sigma^2_{\bar{x}_i}+\tau^2)}}}{\sqrt{2\pi(\sigma^2_{\bar{x}_i}+\tau^2)}} + \frac{1}{N}\sum_{i\in A_1} h_i(x) \tag{12}$$

where $h_i(.)$ is the density function for $\bar{x}_i$ when $i \in A_1$, which is $g_i(\mu)$ convolved with the a normal distribution describing the noise in $\bar{x}_i$: $h_i(x) = \int g_i(\mu)\frac{1}{\sqrt{2\pi\sigma^2_{\bar{x}_i}}}e^{-\frac{(x-\mu)^2}{2\sigma^2_{\bar{x}_i}}}d\mu$.

Like in truncated MLE method, we assume that A_1's contribution to the region $[-\delta_0, \delta_0]$ is negeligiable. Doing a first order Taylor expansion of the marginal density function $f(x|\tau)$ and for $x \in [-\delta_0, \delta_0]$, we have

$$f(x|\tau) \approx \frac{1}{N}\sum_{i\in A_0}\frac{1}{\sqrt{2\pi(\sigma^2_{\bar{x}_i}+\tau^2)}}(1-\frac{x^2}{2(\sigma^2_{\bar{x}_i}+\tau^2)})$$

$$l(x) = \log f(x) \approx C - \sum_{i\in A_0}\frac{x^2}{2(\sigma^2_{\bar{x}_i}+\tau^2)^{\frac{3}{2}}\sum_{i\in A_0}\frac{1}{\sqrt{\sigma^2_{\bar{x}_i}+\tau^2}}}$$

In other words, let $\hat{b}(x)$ be the observation count at the bin centered at x after binning observations in $[-\delta_0, \delta_0]$. We can estimate the coefficient before τ^2 simply by the following,

1. Fit the poisson regression model:

   $$\hat{b}(x) \sim Poisson(\mu(x)); \;\; \log(\mu(x)) = l(x) + \log(N) = \alpha + \beta x^2$$

   and denote the fitted coefficient to x^2 as $\hat{\beta}$.

   (Such binning and fitting steps are also used by the R function **locfdr**.)

2. Use the relationships below and do grid search of τ^2:

   $$\sum_{i=1}^{N}\frac{1}{2(\sigma^2_{\bar{x}_i}+\tau^2)^{\frac{3}{2}}\sum_{i=1}^{N}\frac{1}{\sqrt{\sigma^2_{\bar{x}_i}+\tau^2}}} = -\hat{\beta}$$

   with $\sigma^2_{\bar{x}_i}$ replaced by $\hat{\sigma}^2_{\bar{x}_i}$.

Intuitively, the three procedures are different in several perspectives:

1. ITEB starts by treating the full data set as null and iteratively removing genes with large values, it usually ends up estimating using a set of genes much larger than the other two methods. As a result, it is be able to utilize more information from the data, but it suffers more from initially overestimating τ^2 when γ is large (We care about small γ in our case).

2. CM relies on the first order Taylor expansion of the log likelihood around a small region near 0 if τ^2 is not large enough and needs to plug in variance estimates in the denominator, which makes it less accurate when τ^2 is not large.

3. Both the truncated MLE method and CM need to know the specific form of the likelihood for the null distribution, while ITEB uses only the moments, which makes its application to complicated distributions straightforward.

In Appendix D, we compare performances of the three estimates in different scenarios and discuss their strengths and weaknesses.

D Results with different τ^2-estimation approaches

D.1 Simulation: Estimate of τ^2

For simplicity, we focus on the one-sample setting and generate data under various values of τ and non-null proportion $\gamma = \frac{|A_1|}{N}$. Specifically, we fix $N = 15000, m = 10$, which is of the same order as typical knock-down data. For any given τ and γ, where $\gamma = 0, 2\%, 5\%, 7\%, 10\%$ and $\tau = 0, 0.1, ..., 1, 1.5, 2, 2.5, 3$, we generate the data as below.

1. Generate μ_is: $\mu_i \sim \begin{cases} N(0, \tau^2) & \forall i \in A_0 \\ \pm U[1, \max(3, 10\tau)] & \forall i \in A_1 \end{cases}$

   where $U[1, \max(3, 10\tau)]$ is the uniform distribution between 1 and $\max(3, 10\tau)$, and the signs of μ_is will be half positive and half negative.

2. Generate variances for genes in one of the two settings:

   (a) Independently generate $\sigma_i^2 \sim \chi_1^2$.

   (b) Sample σ_i^2 from its empirical distribution in the real data set, scaled to have mean level 1.

We used the three approaches to estimate τ with $(\alpha_1, \alpha_2) = (0.1, 0.01)$ for ITEB and leave-out proportion to be 0.05 both for truncated MLE and CM. Motivated by the ITEB estimation, we can also set this proportion adaptively. For example, a simplest approach will consist of two-steps: (1) get an initial estimation of τ^2, based on the truncated MLE/CM estimation with a large left-out proportion, say, 0.2, and (2) let J be the set contains the rejected null hypotheses as described in the ITEB algorithm with this initial estimation, we then get the truncated MLE estimation or the CM estimation with the leave-out proportion being $\frac{|J|}{N}$. We also include the adaptive truncated MLE and the adaptive CM estimations in our results, and the oracle estimation where $\hat{\tau}^2 = [\frac{\sum_{i \in A_0}(\bar{x}_i^2 - \hat{\sigma}^2_{\bar{x}_i})}{|A_0|}]_+$ with known null set A_0 as a benchmark for better comparison.

We repeat the simulations 100 times and plot the square root of relative mean ℓ_2 loss err $= \sqrt{\frac{\sum_{i=1}^{100}(\hat{\tau}^2 - \tau^2)^2}{100} \frac{1}{(\tau^2+0.2)}}$ for visualization, results are given in Figure 1. We have also consider the case where the data is not normal by generating parameters in the same way but with Laplacian distributed noise (means and variances are matched), results are given in Figure 2.

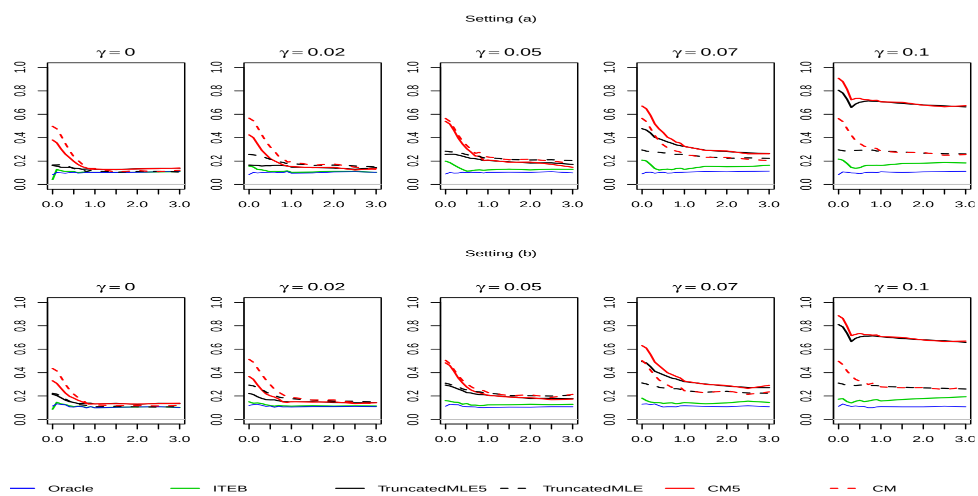

Figure 1: *Gaussian noise, with the upper half being setting (a) and the lower half being setting (b). The curves represent $\sqrt{\text{err}} \sim \tau$. TruncatedMLE and CM represent the two alternative estimation methods with adaptive leave-out proportions while TruncatedMLE5 and CM5 represent the estimations with leave-out proportion being 0.05.*

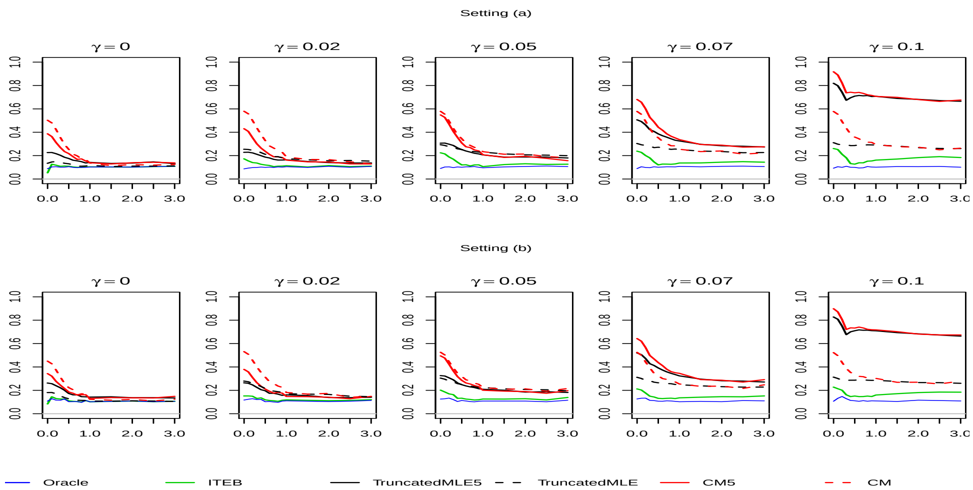

Figure 2: *Laplacian noise, with the upper half being setting (a) and the lower half being setting (b). The curves represent* $\sqrt{\text{err}} \sim \tau$. *TruncatedMLE and CM represent the two alternative estimation methods with adaptive leave-out proportions while TruncatedMLE5 and CM5 represent the estimations with leave-out proportion being 0.05.*

We can see that

1. The two non-adaptive estimations produce much worse estimations when the non-null proportion is large. If we set the leave-out proportion to be large, say, 0.1, although we will suffer less from large γ, we will large variances. The adaptive leave-out proportion is a solution to this problem. Figure 3 shows the square root of the standard deviation sd with leave-out proportion being 0.05 and 0.2.

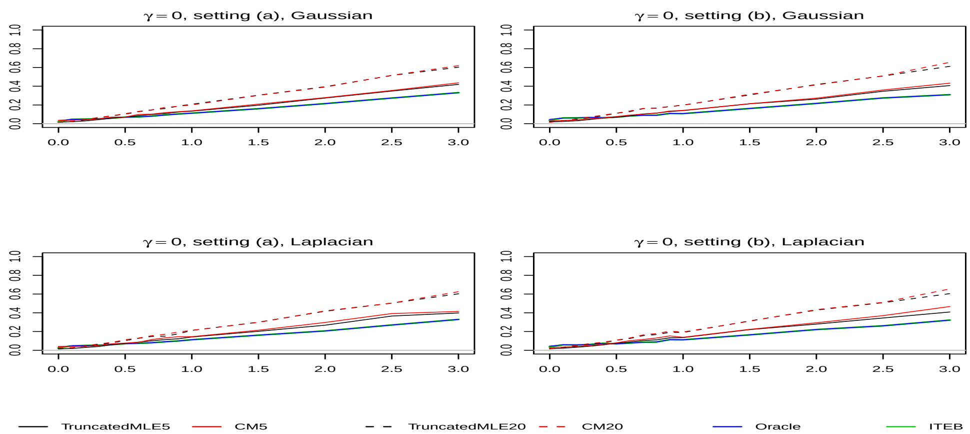

Figure 3: *Laplacian noise, with the upper half being setting (a) and the lower half being setting (b). The curves represent* $\sqrt{\mathrm{sd}} \sim \tau$.

2. CM, adaptive or non-adaptive, is not a good approach estimating τ^2 when τ^2 is not large. It is not surprising as CM is a result of the first order approximation.

3. Both ITEB and the adaptive truncated MLE achieve the adaptive goal successfully and ITEB's performance is as good as the adaptiveTruncated MLE, if not better, across the parameters we have considered. When γ is small, the ITEB estimation is as good as the oracle estimation and it has quite good performance even with γ as large as 0.1.

D.2 Real data results with different τ^2-estimation approaches

In this section, we show the real data results using ITEB, adaptive truncated MLE and adaptive CM. Figure 4 provides quality evaluation of the top K genes in different ranking lists respectively. The evaluation becomes very volatile when the number of selected genes is too small and we add an vertical grey line representing where the top 50 genes is.

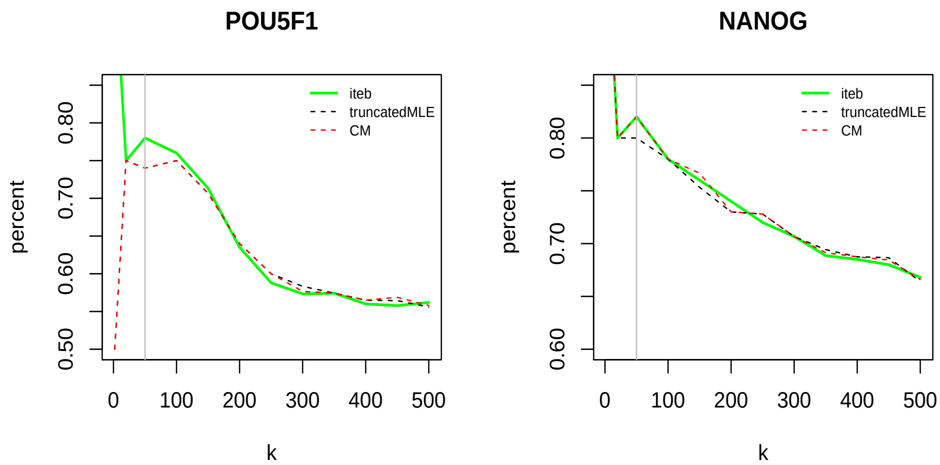

Figure 4: *Percent of genes with ChIP-seq nearby versus versus selected gene size. The x-axis is k, the threshold of the ranking on the whole list created with different τ^2 estimation approaches, and we only consider the top k genes from each ranking list.*

We can see that the ITEB, the truncated MLE and CM result in ranking list that are very similar, and Table 1 shows the overlapping of the top K genes based on the CM/truncated MLE estimation and the ITEB estimation, for $K = 50, 100, 150, 200$.

Table 1: Ranking lists overlapping based on ITEB and other estimation approaches

	POU5F1		NANOG	
K	TruncatedMLE	CM	TruncatedMLE	CM
50	0.84	0.88	0.90	0.92
100	0.93	0.95	0.92	0.96
150	0.95	0.96	0.92	0.93
200	0.94	0.96	0.94	0.95

The cut-offs based on FWER and FDR are more sensitive to the τ^2 estimation, and the ITEB provides more conservative cut-offs.

Table 2: S_0, S_1, S_2 results with FDR/FWER level set at 0.01

	TF	size(S_0)	percent(S_0)	size(S_1)	percent(S_1)	size(S_2)	percent(S_2)
FDR	POU5F1	87	0.73	271	0.61	198	0.66
	NANOG	43	0.81	158	0.77	108	0.78
FWER	POU5F1	31	0.74	85	0.74	70	0.75
	NANOG	20	0.8	50	0.80	41	0.80

The third, fifth and seventh columns are the percent of genes with Chip-seq+Hi-C support in the gene sets based on S_0 (ITEB), S_1 (truncated MLE) and S_2 (CM) respectively.

E Materials and Methods for the knock-down experiment

In this section, we provide more details about how the data is generated and justification of pooling data across days.

Cell Culture: Mouse ES cell line R1 was obtained from Dr. Douglas Melton lab (Harvard University, MA) and cultured under standard conditions. The cells were maintained on gelatin-coated dishes in RPMI knockout medium with 15% knockout serum replacement (KSR), 2 mM L-glutamine, 1 mM non-essential amino acids, 0.55 mM 2-mercaptoethanol (Invitrogen, CA), and 1000 units/mL murine leukaemia inhibitory factor (Chemicon International, CA). Cells were incubated in a 5% CO2–air mixture at 37^{o}C. Cultures were routinely passaged with 0.25% trypsin-EDTA (Invitrogen, CA) and split 1:8 every 2 days. Normal karyotype of ESC was routinely confirmed by analysis of chromosome spreads.

RNA Interference: RNA interference (RNAi) experiments were performed with Nucleofector technology. Briefly, 12 μd of plasmid DNA was transfected into 3.5×10^6 mouse ES cells using the Mouse ES cell Nucleofector kit (Lonza, Switzerland). After nucleofection, the cells were incubated in 500 μl warm ES medium for 15 min. Then, the cells were split into four gelatin-coated 60-mm tissue culture plates containing 5 ml of warm ES medium. Puromycin selection was introduced 18 h later at 1 μ g/ml, and the medium was changed daily. 30 h, 48 h, and 72 h after puromycin selection, the cells were collected for RNA isolation.

Microarray and Data Processing: Microarray hybridizations were performed on the MouseRef-8 v2.0 expression beadchip arrays (Illumina, CA). To prepare sample, 200 ng of total RNA was reverse transcribed, followed by a T7 RNA polymerase-based linear amplification using the Illumina TotalPrep RNA Amplification kit (Applied Biosystems, CA).

After amplification, 750 ng of biotin-labeled cRNA was hybridized to gene specific probes attached to the beads, and the expression levels of transcripts were measured simultaneously.

References

Gordon, R. D. (1941). Values of mills' ratio of area to bounding ordinate and of the normal probability integral for large values of the argument. *The Annals of Mathematical Statistics 12*(3), 364–366.

Soms, A. P. (1976). An asymptotic expansion for the tail area of the t-distribution. *Journal of the American Statistical Association 71*(355), 728–730.