kernelPSI: a Post-Selection Inference Framework for Nonlinear Variable Selection Supplementary Materials

A Proof of Lemma 1

In this Appendix, for a linear kernel of the outcome $L = YY^T$, we detail the necessary steps to transform the empirical HSIC estimators into a quadratic form. For the biased estimator, the result is straightforward:

$$\begin{aligned}\widehat{\mathrm{HSIC}}_{\text{biased}} &= \frac{1}{(n-1)^2}\ \mathrm{trace}(K\Pi_n L \Pi_n) \\ &= \frac{1}{(n-1)^2} Y^T(\Pi_n K \Pi_n) Y \\ &= Y^T Q_{\text{biased}} Y,\end{aligned}$$

where $\Pi_n = I_{n\times n} - \frac{1}{n} 1_n {1_n}^T$.

For the unbiased estimator, the calculations are more tedious:

$$\widehat{\mathrm{HSIC}}_{\text{unbiased}}(K,L) = \frac{1}{n(n-3)}\left[\mathrm{trace}(\underline{K}\ \underline{L}) + \frac{1^T\underline{K}1\ 1^T\underline{L}1}{(n-1)(n-2)} - \frac{2}{n-2} 1^T \underline{K}\ \underline{L}1\right],$$

where $\underline{K} = K - \mathrm{diag}(K)$ and $\underline{L} = L - \mathrm{diag}(L)$. The diagonal matrices $\mathrm{diag}(K)$ and $\mathrm{diag}(L)$ can be respectively rewritten as : $\mathrm{diag}(K) = \sum_{i=1}^n P^{(i)} K P^{(i)}$ and $\mathrm{diag}(L) = \sum_{i=1}^n P^{(i)} L P^{(i)}$. $P^{(i)}$ is the projection on the i^{th} coordinate. We remark that $P^{(i)}P^{(j)} = \delta_{ij} P^{(i)}$.

We now develop each term of the previous equation.

$$
\begin{aligned}
\operatorname{trace}(\underline{K}\ \underline{L}) &= \operatorname{trace}\left(KL - \sum_{i=1}^{n} KP^{(i)}LP^{(i)} - \sum_{i=1}^{n} P^{(i)}KP^{(i)}L + \sum_{i,j=1}^{n} P^{(j)}KP^{(j)}P^{(i)}LP^{(i)}\right) \\
&= \operatorname{trace}(KL) - \operatorname{trace}\left(\sum_{i=1}^{n} P^{(i)}KP^{(i)}L\right) - \operatorname{trace}\left(\sum_{i=1}^{n} P^{(i)}KP^{(i)}L\right) \\
&+ \operatorname{trace}\left(\sum_{i,j=1}^{n} P^{(j)}KP^{(j)}P^{(i)}LP^{(i)}\right) \\
&= \operatorname{trace}(KL) - 2\ \operatorname{trace}\left(\sum_{i=1}^{n} P^{(i)}KP^{(i)}L\right) + \operatorname{trace}\left(\sum_{i}^{n} P^{(i)}KP^{(i)}LP^{(i)}\right) \\
&= \operatorname{trace}(KL) - \operatorname{trace}\left(\sum_{i=1}^{n} P^{(i)}KP^{(i)}L\right) \\
&= Y^TKY - Y^T\left(\sum_{i=1}^{n} P^{(i)}KP^{(i)}\right)Y \\
&= Y^T(K - K_P)Y \qquad \text{with} \qquad K_P = \sum_{i=1}^{n} P^{(i)}KP^{(i)} \\
\operatorname{trace}(\underline{K}\ \underline{L}) &= Y^TK_1K
\end{aligned}
$$

Similarly, we obtain:

$$
\begin{aligned}
1^T\underline{K}1 &= 1^T\left(K - K_P\right)1 = c_X \\
1^T\underline{L}1 &= 1^T\left(L - \sum_{i=1}^{n} P^{(i)}LP^{(i)}\right)1 \\
&= Y^T11^TY\ - \operatorname{trace}\left(11^T\sum_{i=1}^{n} P^{(i)}LP^{(i)}\right) \\
&= Y^T\left(11^T - \sum_{i=1}^{n} P^{(i)}11^TP^{(i)}\right)Y \\
1^T\underline{L}1 &= Y^TK_2Y
\end{aligned}
$$

As for the last term:

$$
\begin{aligned}
1^T \underline{K}\ \underline{L} 1 &= \operatorname{trace}(11^T \underline{K}\ \underline{L}) \\
&= \operatorname{trace}\left(11^T \left(K - \sum_{i=1}^{n} P^{(i)} K P^{(i)}\right)\ \left(L - \sum_{j=1}^{n} P^{(j)} L P^{(j)}\right)\right) \\
&= \operatorname{trace}(11^T K\ L) - \operatorname{trace}\left(11^T K \sum_{i=1}^{n} P^{(i)} L P^{(i)}\right) - \operatorname{trace}\left(11^T \sum_{i=1}^{n} P^{(i)} K P^{(i)} L\right) \\
&+ \operatorname{trace}\left(11^T \sum_{i=1}^{n} P^{(i)} K P^{(i)} \sum_{j=1}^{n} P^{(j)} L P^{(j)}\right) \\
&= \operatorname{trace}(11^T K\ L) - \operatorname{trace}\left(\sum_{i=1}^{n} P^{(i)} 11^T K P^{(i)} L\right) - \operatorname{trace}\left(\sum_{i=1}^{n} 11^T P^{(i)} K P^{(i)} L\right) \\
&+ \operatorname{trace}\left(11^T \sum_{i=1}^{n} P^{(i)} K P^{(i)} L P^{(i)}\right) \\
&= Y^T (11^T K) Y - Y^T \left(\sum_{i=1}^{n} P^{(i)} 11^T K P^{(i)}\right) Y - Y^T \left(\sum_{i=1}^{n} 11^T P^{(i)} K P^{(i)}\right) Y \\
&+ Y^T \left(\sum_{i=1}^{n} P^{(i)} 11^T P^{(i)} K P^{(i)}\right) Y \\
1^T \underline{K}\ \underline{L} 1 &= Y^T K_3 Y
\end{aligned}
$$

That yields the following quadratic form:

$$
\begin{aligned}
\widehat{\text{HSIC}}_{\text{unbiased}}(X, Y) &= \frac{1}{n(n-3)} \left[Y^T K_1 Y + c_X \frac{Y^T K_2 Y}{(n-1)(n-3)} - \frac{2}{n-2} Y^T K_3 Y \right] \\
&= Y^T Q_{\text{unbiased}} Y
\end{aligned}
$$

B Proof of Theorem 1

For a quadratic kernel association score $s(K, Y) = Y^T Q(K)Y$, we represent the three kernel selection strategies as an intersection of quadratic constraints.

For marginal screening, we can write the selection event of the top S' kernels $i_1, \cdots, i_{S'}$ in the following way :

$$E_{S'}^{\text{screening}} = \bigcap_{l=1}^{S'-1} \left\{ Y^T Q(K_{i_l})Y \geq Y^T Q(K_{i_{l+1}})Y \right\} \cap \bigcap_{l' \notin \{i_1, \cdots, i_{S'}\}} \left\{ Y^T Q(K_{i_{S'}})Y \geq Y^T Q(K_{l'})Y \right\}$$

In Yamada et al. [1], the authors obtain $S'(S - S')$ constraints by comparing the association score of each selected kernel to the association scores of all discarded kernels. Here, by conditioning on the order of selection of the kernels, we only obtain $S - 1$ constraints in $E_{S'}^{\text{screening}}$.

For forward stepwise selection (Algorithm 1), we first start by modeling an intermediate step s. The selection of the kernel K_{i_s} is equivalent to the following selection event:

$$\bigcap_{\substack{i \notin \mathcal{J}^{(s-1)} \\ i \neq i_s}} \left\{ Y \text{ s.t. } Y^T Q(K_{\mathcal{J}^{(s-1)} \cup \{i_s\}})Y \geq Y^T Q(K_{\mathcal{J}^{(s-1)} \cup \{i\}})Y \right\},$$

where $\mathcal{J}^{(m)}$ represents the set of selected kernels at step m and $K_{\mathcal{A}} = \sum_{p \in \mathcal{A}} K_p$ for a subset $\mathcal{A}$ of $\{1, \cdots, S\}$.

We can then recursively define the event E_s, representing the selection of $s \leq S'$ groups:

$$E_s^{\text{forward}} = E_{s-1}^{\text{forward}} \cap \bigcap_{\substack{i \notin \mathcal{J}^{(s-1)} \\ i \neq i_s}} \left\{ Y \text{ s.t. } Y^T Q(K_{\mathcal{J}^{(s-1)} \cup \{i_s\}})Y \geq Y^T Q(K_{\mathcal{J}^{(s-1)} \cup \{i\}})Y \right\}$$

For $s = S'$, we then obtain a conjunction of quadratic constraints.

For backward selection (Algorithm 2), we can derive a similar set of recursive constraints to model the elimination of the kernels $\mathcal{I}^{(s)} = \{i_1, \cdots, i_s\}$:

$$E_s^{\text{backward}} = E_{s-1}^{\text{backward}} \cap \bigcap_{\substack{i \notin \mathcal{I}^{(s-1)} \\ i \neq i_s}} \left\{ Y \text{ s.t. } Y^T Q(K^-_{\mathcal{I}^{(s-1)} \cup \{i_s\}})Y \geq Y^T Q(K^-_{\mathcal{I}^{(s-1)} \cup \{i\}})Y \right\},$$

where $K^-_{\mathcal{A}} = \sum_{p \in \mathcal{A}^c} K_p$. The set $\mathcal{A}^c$ is the complement of $\mathcal{A}$ in $\{1, \cdots, S\}$.

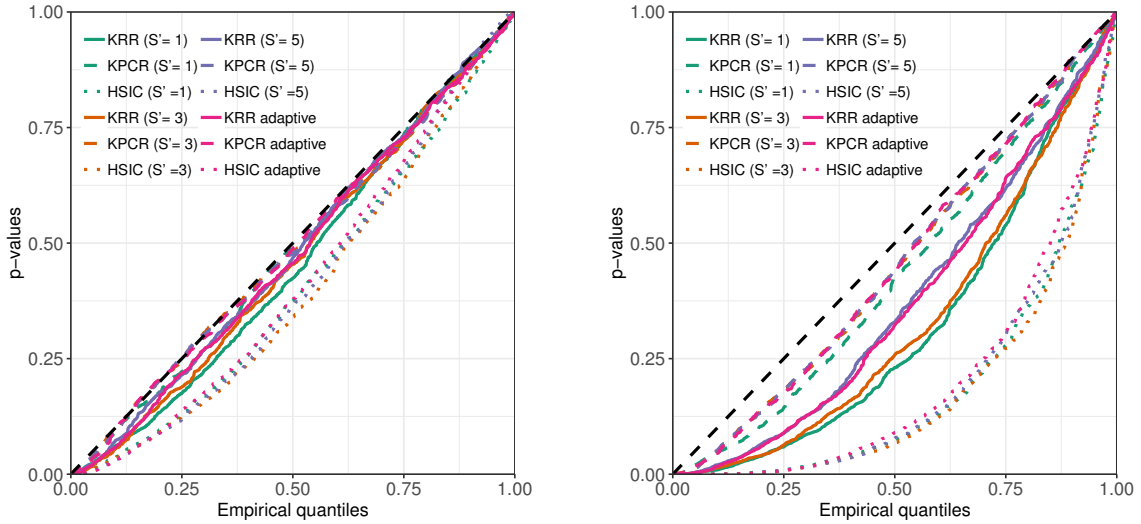
To model the a posteriori choice of S' in the adaptive variants, an additional set of constraints must be introduced in the selection event. In Equation 1, we model the selection event $E_{\text{adaptive}}^{\text{forward}}$ corresponding to the adaptive extension of forward stepwise selection. The quadratic set of constraints in E_S^{forward} represents the order of selection of the kernels $\mathcal{J}^{(S)} = \{i_1, \cdots, i_S\}$, while the intersection of the other constraints represents the selection of the number of kernels S'. The backward version $E_{\text{adaptive}}^{\text{backward}}$ can be easily deduced in a similar fashion.

$$E_{\text{adaptive}}^{\text{forward}} = E_S^{\text{forward}} \cap \bigcap_{\substack{m=1 \\ m \neq S'}}^{S} \left\{ Y \text{ s.t. } Y^T Q(K_{\mathcal{J}^{(S')}})Y \geq Y^T Q(K_{\mathcal{J}^{(m)}})Y \right\} \tag{1}$$

The result in Theorem 1 is more general by adding to the quadratic form a constant, which can be used as a form of penalization. The above proof can be easily extended to the setting of Theorem 1.

C Experiments

C.1 Statistical validity: Statistical power of kernelPSI for different effect sizes, on simulated data



(a) $\theta = 0.1$

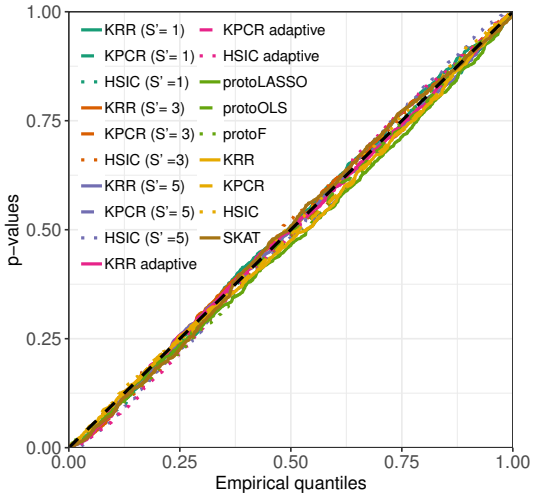
(b) $\theta = 0.2$

(c) $\theta = 0.4$

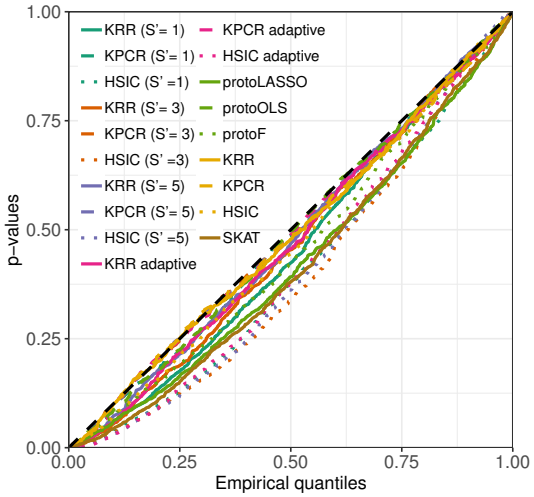
(d) $\theta = 0.5$

Figure 1: Q-Q plots comparing the empirical kernelPSI p-values distributions under the alternative hypothesis to the uniform distribution, for different effect sizes θ. The data is generated as described in Section 7.1.

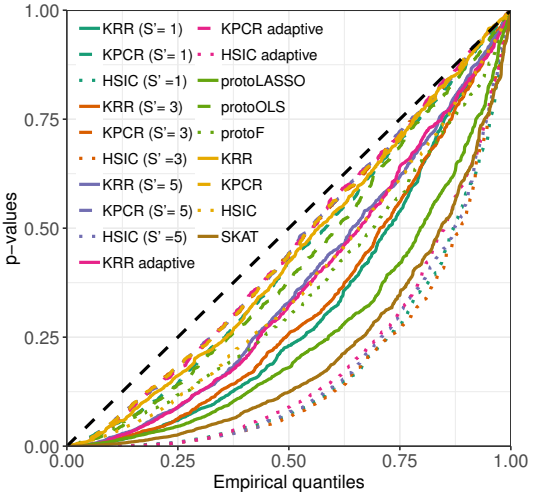
C.2 Benchmarking for the first configuration: using Gaussian kernels over simulated Gaussian data



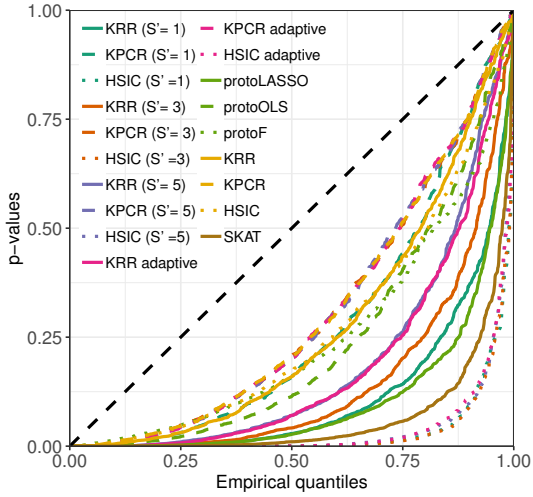
(a) $\theta = 0.0$



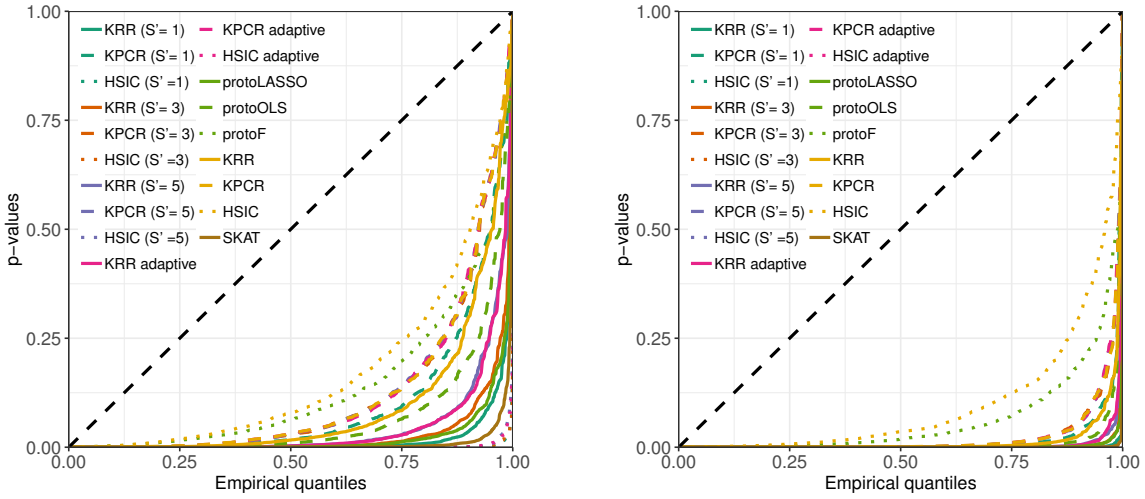
(b) $\theta = 0.1$



(c) $\theta = 0.2$



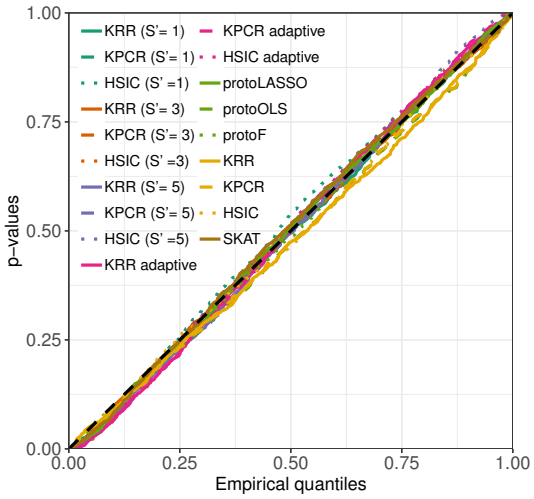
(d) $\theta = 0.3$



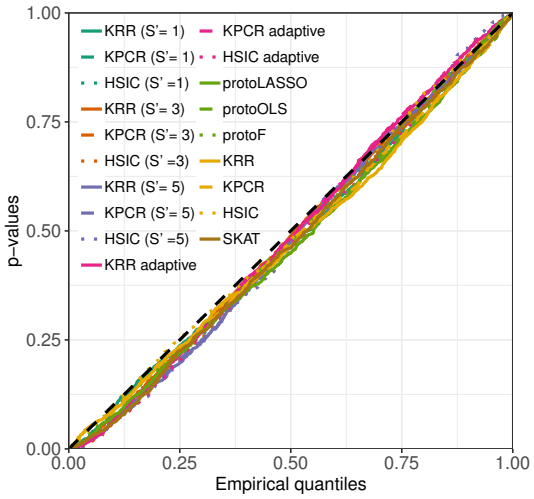
(e) $\theta = 0.4$ (f) $\theta = 0.5$

Figure 2: Q-Q plots comparing the empirical kernelPSI and benchmarking p-values distributions under the null ($\theta = 0$) or alternative hypothesis ($\theta > 0$) to the uniform distribution, for different effect sizes θ, using Gaussian kernels for simulated Gaussian data. The data generation and benchmarked methods are described in Section 7.2.

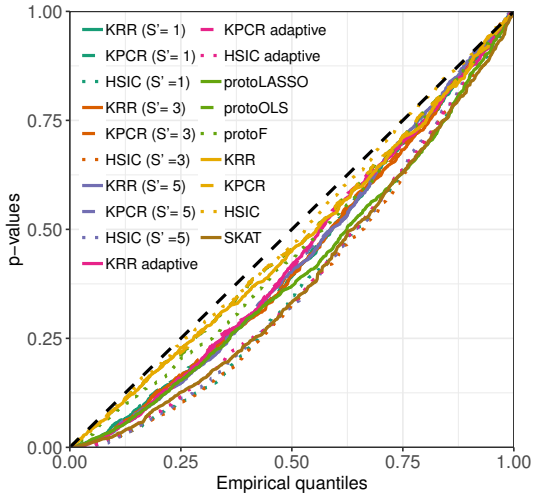
C.3 Benchmarking for the second configuration: using linear kernels over simulated binary data



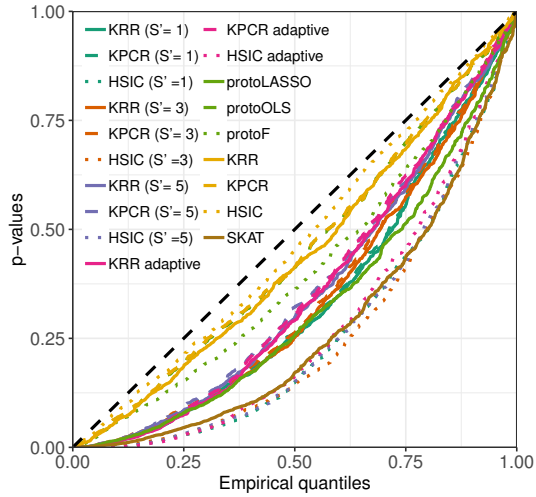
(a) $\theta = 0.0$



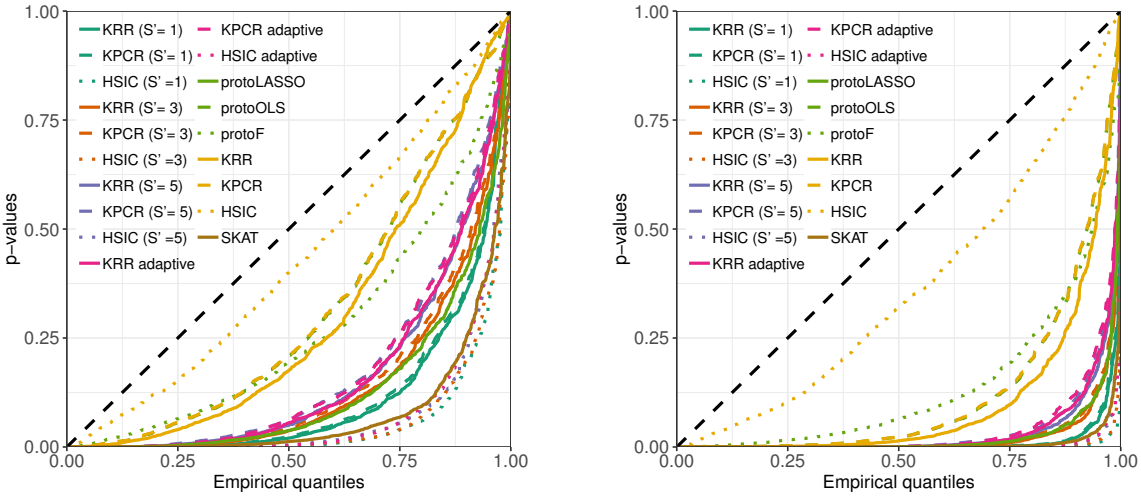
(b) $\theta = 0.01$



(c) $\theta = 0.02$

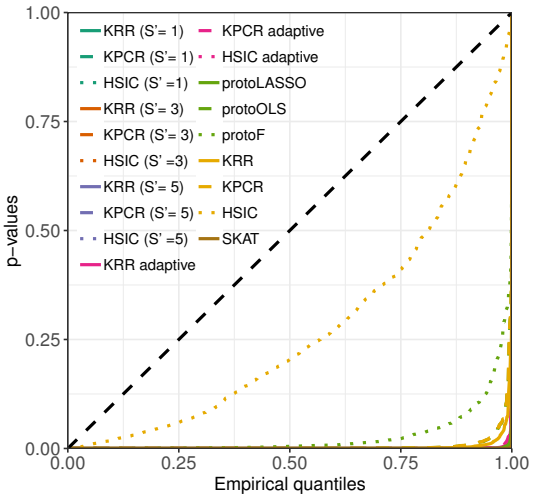


(d) $\theta = 0.03$



(a) $\theta = 0.05$

(b) $\theta = 0.07$

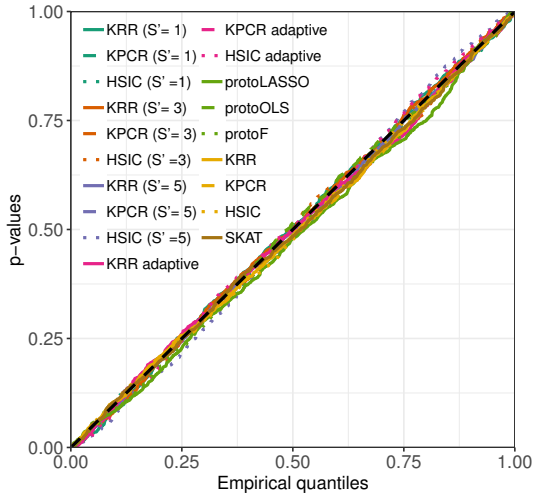


(c) $\theta = 0.10$

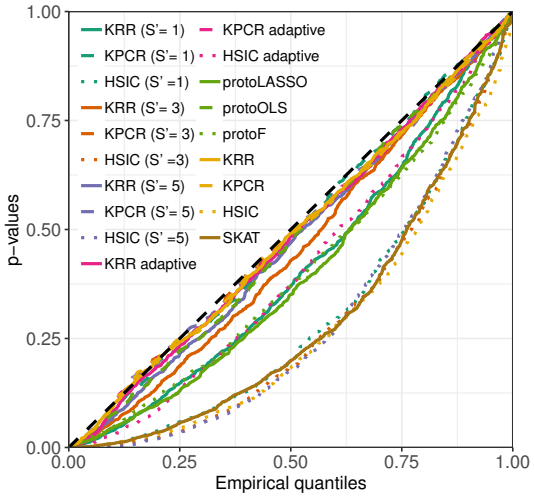
Figure 4: Q-Q plots comparing the empirical kernelPSI and benchmarking p-values distributions under the null ($\theta = 0$) or alternative hypothesis ($\theta > 0$) to the uniform distribution, for different effect sizes θ, using linear kernels for simulated binary data. The data generation and benchmarked methods are described in Section 7.2.

C.4 Benchmarking for the third configuration: using Gaussian kernels over simulated Swiss roll data

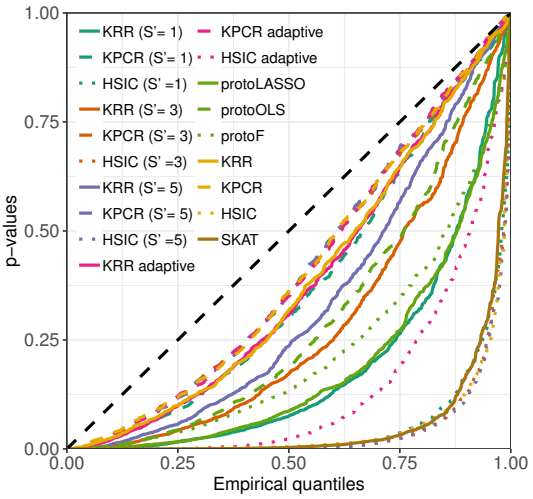
C.4.1 Statistical validity: Q-Q plots for various effect sizes



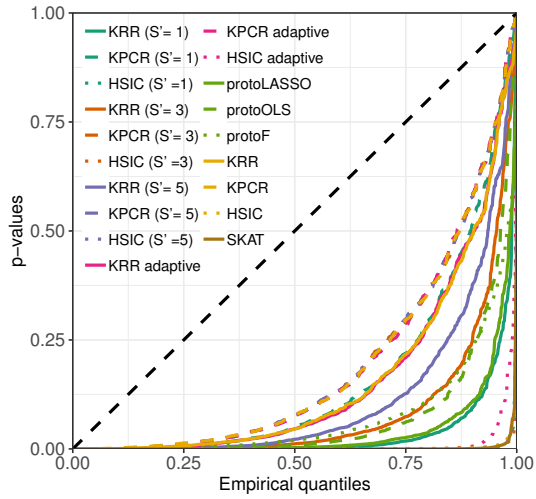
(a) $\theta = 0.0$



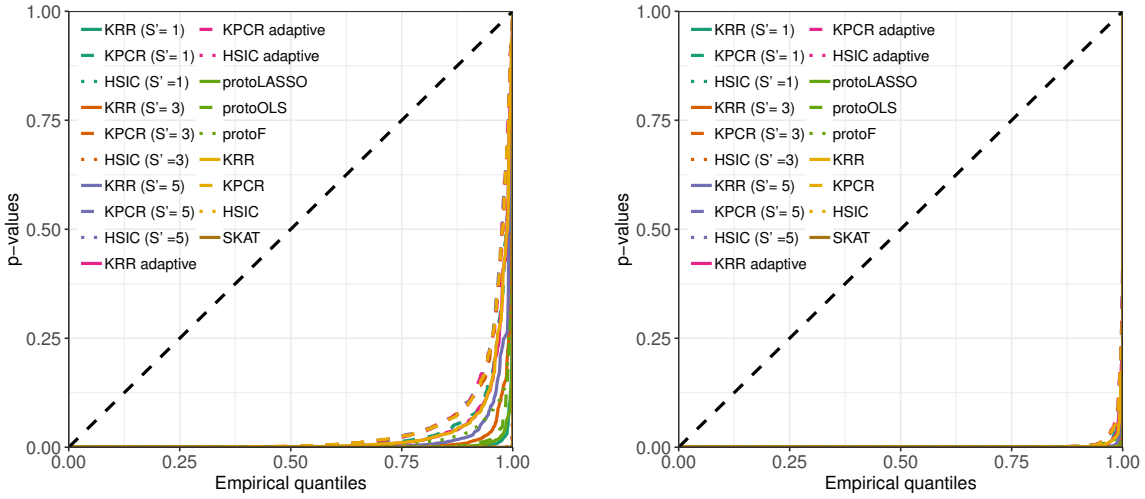
(b) $\theta = 0.1$



(c) $\theta = 0.2$



(d) $\theta = 0.3$



(e) $\theta = 0.4$ (f) $\theta = 0.5$

Figure 5: Q-Q plots comparing the empirical kernelPSI and benchmarking p-values distributions under the null ($\theta = 0$) or alternative hypothesis ($\theta > 0$) to the uniform distribution, for different effect sizes θ, using Gaussian kernels for simulated Swiss roll data. The data generation and benchmarked methods are described in Section 7.2.

C.4.2 Evolution of the statistical power as a function of the effect size

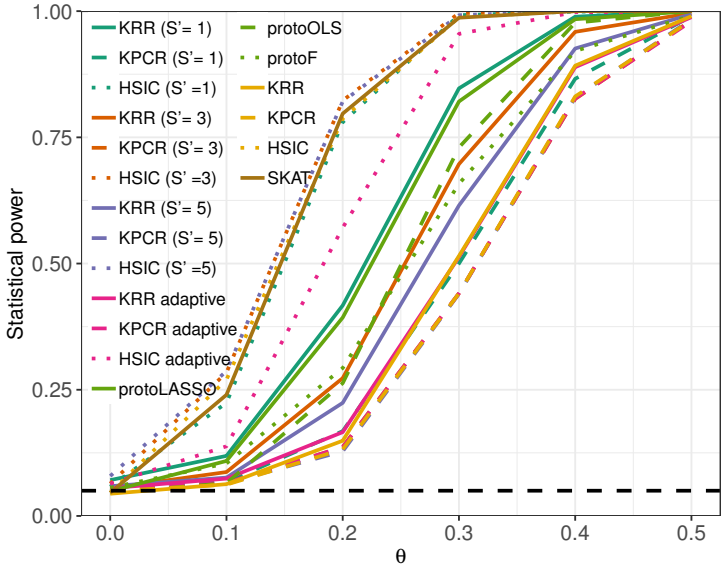


Figure 6: Statistical power of kernelPSI variants and benchmark methods, using Gaussian kernels for simulated Swiss roll data.

C.5 Kernel selection performance

Table 1: Ability of the kernel selection procedure to recover the true causal kernels, using linear kernels over with binary data.

	$\boldsymbol{\theta}$	$S' = 1$	$S' = 3$	$S' = 5$	Adaptive
Recall	0.0	0.102	0.316	0.529	0.390
	0.01	0.117	0.341	0.553	0.414
	0.02	0.163	0.388	0.590	0.482
	0.03	0.224	0.445	0.632	0.559
	0.05	0.311	0.512	0.683	0.639
	0.07	0.332	0.542	0.713	0.686
	0.1	0.333	0.581	0.759	0.750
Precision	0.0	0.308	0.316	0.317	0.320
	0.01	0.351	0.341	0.331	0.335
	0.02	0.489	0.388	0.354	0.377
	0.03	0.673	0.445	0.379	0.422
	0.05	0.935	0.512	0.409	0.463
	0.07	0.996	0.542	0.428	0.473
	0.1	1.000	0.581	0.455	0.484

Table 2: Ability of the kernel selection procedure to recover the true causal kernels, using Gaussian kernels over simulated Swiss roll data.

	$\boldsymbol{\theta}$	$S' = 1$	$S' = 3$	$S' = 5$	Adaptive
Recall	0.0	0.112	0.306	0.505	0.488
	0.1	0.125	0.331	0.541	0.715
	0.2	0.157	0.404	0.579	0.963
	0.3	0.196	0.462	0.621	0.999
	0.4	0.239	0.507	0.645	1.000
	0.5	0.275	0.537	0.655	1.000
Precision	0.0	0.337	0.306	0.303	0.328
	0.1	0.377	0.331	0.325	0.318
	0.2	0.471	0.404	0.347	0.303
	0.3	0.588	0.462	0.372	0.300
	0.4	0.717	0.507	0.387	0.300
	0.5	0.825	0.537	0.393	0.300

C.6 *A. thaliana* case study: data description and pre-processing

For this dataset, we are interested in the effect of each gene on the outcome Y, which corresponds to the flowering time in green house, corrected for population structure. We follow the same correction procedure as in Azencott et al. [2]. The total number of samples is $n = 166$. The features are $9\,938$ binary SNPs located within a ± 20-kilobase window of 174 pre-selected genes. These genes, known as candidate genes, have been selected by experts as most likely to be involved in flowering time traits. The full list of genes with additional functional information is available from the following URL: `https://www.mpipz.mpg.de/14637/Arabidopsis_flowering_genes`.

We start with applying hierarchical clustering algorithm to define clusters within each gene. For a given cluster, the associated SNPs are expected to be in linkage disequilibrium. The genes are clustered differently depending on the sample size. Genes with a number of SNPs lower than the gene median size (58 SNPs) are split into 6 clusters. We apply the fixed version of kernelPSI for the three parameterizations $S' \in \{1, 2, 4\}$. For genes larger than the median size, we split them into 12 clusters and consider a number of selected clusters $S' \in \{1, 3, 6\}$.

We use the identical-by-state (IBS) kernel [3] for the clusters. This kernel is commonly used in GWAS. For two samples i and j, the IBS kernel corresponds to the fraction of identical SNPs between the two samples:

$$K_{ij} = \frac{|X_i| - ||X_i - X_j||}{|X_i|},$$

where $|X_i|$ is the length of X_i.

C.7 *A. thaliana* case study: rank concordance between the methods

Table 3: Concordance between kernelPSI and benchmark methods, measured by the Kendall's tau coefficient between the p-values returned for the 50% smallest genes.

	KRR (S'= 1)	KPCR (S'= 1)	HSIC (S'= 1)	KRR (S'= 2)	KPCR (S'= 2)	HSIC (S'= 2)	KRR (S'= 4)	KPCR (S'= 4)	HSIC (S'= 4)	KRR adaptive	KPCR adaptive	HSIC adaptive	protoLASSO	protoOLS	KRR	KPCR	HSIC	SKAT
KRR (S'= 1)	1.000	0.795	0.664	0.592	0.501	0.593	0.359	0.204	0.443	0.556	0.412	0.410	0.497	0.198	0.263	0.174	0.421	0.424
KPCR (S'= 1)	0.795	1.000	0.579	0.517	0.566	0.495	0.306	0.258	0.327	0.512	0.486	0.368	0.380	0.230	0.273	0.218	0.362	0.369
HSIC (S'= 1)	0.664	0.579	1.000	0.467	0.361	0.655	0.217	0.071	0.552	0.393	0.251	0.529	0.481	0.060	0.150	0.055	0.575	0.577
KRR (S'= 2)	0.592	0.517	0.467	1.000	0.724	0.553	0.491	0.293	0.453	0.580	0.434	0.367	0.495	0.280	0.363	0.255	0.418	0.424
KPCR (S'= 2)	0.501	0.566	0.361	0.724	1.000	0.396	0.410	0.326	0.273	0.533	0.507	0.284	0.336	0.250	0.325	0.229	0.311	0.315
HSIC (S'= 2)	0.593	0.495	0.655	0.553	0.396	1.000	0.255	0.159	0.664	0.441	0.323	0.561	0.504	0.195	0.214	0.181	0.544	0.549
KRR (S'= 4)	0.359	0.306	0.217	0.491	0.410	0.255	1.000	0.621	0.272	0.486	0.394	0.182	0.324	0.498	0.666	0.494	0.181	0.179
KPCR (S'= 4)	0.204	0.258	0.071	0.293	0.326	0.159	0.621	1.000	0.115	0.346	0.460	0.081	0.155	0.581	0.578	0.620	0.070	0.069
HSIC (S'= 4)	0.443	0.327	0.552	0.453	0.273	0.664	0.272	0.115	1.000	0.360	0.214	0.475	0.471	0.169	0.183	0.124	0.569	0.571
KRR adaptive	0.556	0.512	0.393	0.580	0.533	0.441	0.486	0.346	0.360	1.000	0.685	0.403	0.429	0.330	0.438	0.322	0.307	0.307
KPCR adaptive	0.412	0.486	0.251	0.434	0.507	0.323	0.394	0.460	0.214	0.685	1.000	0.241	0.300	0.415	0.419	0.424	0.169	0.176
HSIC adaptive	0.410	0.368	0.529	0.367	0.284	0.561	0.182	0.081	0.475	0.403	0.241	1.000	0.310	0.132	0.144	0.114	0.381	0.385
protoLASSO	0.497	0.380	0.481	0.495	0.336	0.504	0.324	0.155	0.471	0.429	0.300	0.310	1.000	0.218	0.274	0.189	0.483	0.486
protoOLS	0.198	0.230	0.060	0.280	0.250	0.195	0.498	0.581	0.169	0.330	0.415	0.132	0.218	1.000	0.622	0.856	0.107	0.106
KRR	0.263	0.273	0.150	0.363	0.325	0.214	0.666	0.578	0.183	0.438	0.419	0.144	0.274	0.622	1.000	0.641	0.168	0.164
KPCR	0.174	0.218	0.055	0.255	0.229	0.181	0.494	0.620	0.124	0.322	0.424	0.114	0.189	0.856	0.641	1.000	0.092	0.088
HSIC	0.421	0.362	0.575	0.418	0.311	0.544	0.181	0.070	0.569	0.307	0.169	0.381	0.483	0.107	0.168	0.092	1.000	0.972
SKAT	0.424	0.369	0.577	0.424	0.315	0.549	0.179	0.069	0.571	0.307	0.176	0.385	0.486	0.106	0.164	0.088	0.972	1.000

Table 4: Concordance between kernelPSI and benchmark methods, measured by the Kendall's tau coefficient between the p-values returned for the 50% largest genes.

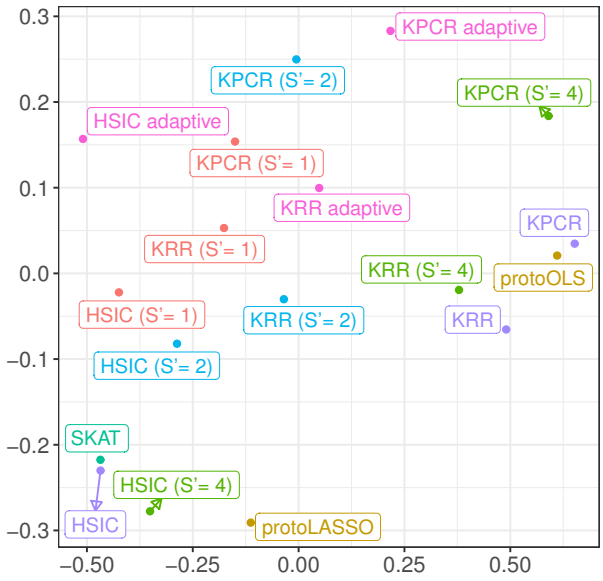
	KRR (S'= 1)	KPCR (S'= 1)	HSIC (S'= 1)	KRR (S'= 3)	KPCR (S'= 3)	HSIC (S'= 3)	KRR (S'= 6)	KPCR (S'= 6)	HSIC (S'= 6)	KRR adaptive	KPCR adaptive	HSIC adaptive	protoLASSO	protoOLS	KRR	KPCR	HSIC	SKAT
KRR (S'= 1)	1.000	0.936	0.778	0.454	0.385	0.450	0.328	0.121	0.385	0.301	0.187	0.332	0.357	-0.032	0.220	0.033	0.522	0.520
KPCR (S'= 1)	0.936	1.000	0.724	0.439	0.396	0.414	0.308	0.129	0.357	0.303	0.189	0.310	0.324	-0.029	0.219	0.034	0.488	0.490
HSIC (S'= 1)	0.778	0.724	1.000	0.421	0.345	0.527	0.366	0.128	0.437	0.299	0.204	0.338	0.397	-0.004	0.202	0.063	0.538	0.536
KRR (S'= 3)	0.454	0.439	0.421	1.000	0.779	0.550	0.495	0.247	0.439	0.403	0.270	0.293	0.380	0.039	0.274	0.105	0.475	0.476
KPCR (S'= 3)	0.385	0.396	0.345	0.779	1.000	0.444	0.449	0.293	0.380	0.381	0.283	0.235	0.282	0.062	0.289	0.149	0.382	0.383
HSIC (S'= 3)	0.450	0.414	0.527	0.550	0.444	1.000	0.327	0.004	0.492	0.273	0.150	0.288	0.354	-0.093	0.133	-0.024	0.516	0.517
KRR (S'= 6)	0.328	0.308	0.366	0.495	0.449	0.327	1.000	0.422	0.411	0.394	0.283	0.270	0.342	0.180	0.442	0.250	0.339	0.336
KPCR (S'= 6)	0.121	0.129	0.128	0.247	0.293	0.004	0.422	1.000	0.127	0.265	0.415	0.069	0.115	0.416	0.350	0.468	0.093	0.094
HSIC (S'= 6)	0.385	0.357	0.437	0.439	0.380	0.492	0.411	0.127	1.000	0.299	0.177	0.350	0.290	0.043	0.205	0.114	0.411	0.405
KRR adaptive	0.301	0.303	0.299	0.403	0.381	0.273	0.394	0.265	0.299	1.000	0.603	0.392	0.345	0.081	0.270	0.156	0.223	0.228
KPCR adaptive	0.187	0.189	0.204	0.270	0.283	0.150	0.283	0.415	0.177	0.603	1.000	0.286	0.248	0.243	0.236	0.327	0.125	0.126
HSIC adaptive	0.332	0.310	0.338	0.293	0.235	0.288	0.270	0.069	0.350	0.392	0.286	1.000	0.250	-0.044	0.163	0.036	0.322	0.321
protoLASSO	0.357	0.324	0.397	0.380	0.282	0.354	0.342	0.115	0.290	0.345	0.248	0.250	1.000	0.084	0.222	0.057	0.345	0.347
protoOLS	-0.032	-0.029	-0.004	0.039	0.062	-0.093	0.180	0.416	0.043	0.081	0.243	-0.044	0.084	1.000	0.332	0.594	-0.037	-0.038
KRR	0.220	0.219	0.202	0.274	0.289	0.133	0.442	0.350	0.205	0.270	0.236	0.163	0.222	0.332	1.000	0.387	0.204	0.205
KPCR	0.033	0.034	0.063	0.105	0.149	-0.024	0.250	0.468	0.114	0.156	0.327	0.036	0.057	0.594	0.387	1.000	-0.001	-0.007
HSIC	0.522	0.488	0.538	0.475	0.382	0.516	0.339	0.093	0.411	0.223	0.125	0.322	0.345	-0.037	0.204	-0.001	1.000	0.984
SKAT	0.520	0.490	0.536	0.476	0.383	0.517	0.336	0.094	0.405	0.228	0.126	0.321	0.347	-0.038	0.205	-0.007	0.984	1.000

C.8 *A. thaliana* case study: list of significant genes

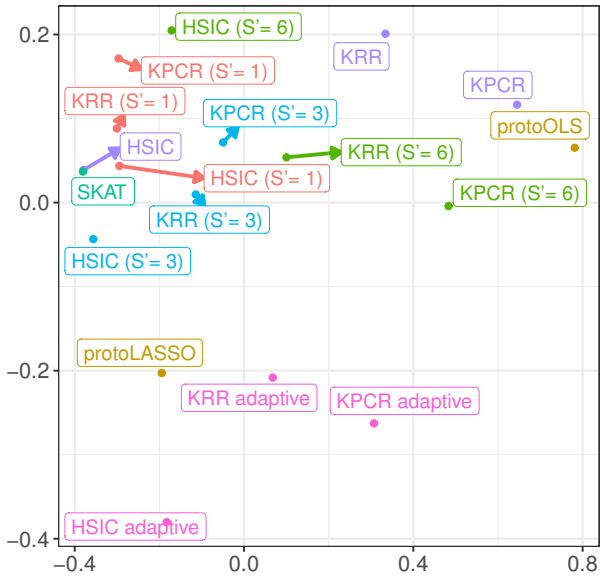
Table 5: Genes detected as significantly associated to the FT GH phenotype, by method.

Method	Significant genes
HSIC (S'= 4)	AT1G53090, AT1G53160, AT1G56170, AT3G60250
KPCR (S'= 1)	AT5G57360, AT4G00650
protoLASSO	–
KPCR (S'= 3)	AT1G80340, AT4G00650
KPCR	–
HSIC (S'= 3)	–
HSIC (S'= 1)	AT5G57360]
KPCR adaptive	AT4G35900, AT5G47640, AT5G55835
HSIC	–
HSIC adaptive	AT1G53160, AT2G18790, AT4G08920, AT5G46210, AT5G47640, AT5G55835
KPCR (S'= 4)	AT2G22540, AT4G35900, AT5G60100
KRR	–
KPCR (S'= 6)	AT1G69120, AT5G10945
KPCR (S'= 2)	AT1G56170
KRR adaptive	AT2G18790, AT2G25930, AT4G00650, AT5G47640, AT5G55835
SKAT	–
KRR (S'= 6)	AT1G69120, AT2G21070
KRR (S'= 4)	AT1G68840, AT4G35900, AT5G60100, AT5G65050, AT5G65070
KRR (S'= 3)	AT1G80340, AT2G21070
KRR (S'= 2)	AT1G56170, AT2G38880, AT5G65060
HSIC (S'= 6)	AT4G08920, AT5G26147, AT5G47640
KRR (S'= 1)	AT5G57360, AT5G65060
HSIC (S'= 2)	AT1G53090, AT1G56170, AT2G27990
protoOLS	–

C.9 *A. thaliana* case study: non-metric multi-dimensional scaling of the results.



(a) NMDS results for the 50% smallest genes



(b) NMDS results for the 50% largest genes

Figure 7: Non-metric multi-dimensional scaling (NMDS) of the p-values obtained by the kernelPSI and benchmark methods on *Arabidopsis thaliana* data, using $1 - \tau$ as a distance.

References

[1] Makoto Yamada, Yuta Umezu, Kenji Fukumizu, and Ichiro Takeuchi. Post selection inference with kernels. In Amos Storkey and Fernando Perez-Cruz, editors, *Proceedings of the Twenty-First International Conference on Artificial Intelligence and Statistics*, volume 84 of *Proceedings of Machine Learning Research*, pages 152–160, Playa Blanca, Lanzarote, Canary Islands, 09–11 Apr 2018. PMLR.

[2] Chloé-Agathe Azencott, Dominik Grimm, Mahito Sugiyama, Yoshinobu Kawahara, and Karsten M. Borgwardt. Efficient network-guided multi-locus association mapping with graph cuts. *Bioinformatics*, 29(13):i171–i179, 2013.

[3] Lydia Coulter Kwee, Dawei Liu, Xihong Lin, Debashis Ghosh, and Michael P. Epstein. A powerful and flexible multilocus association test for quantitative traits. *The American Journal of Human Genetics*, 82(2):386–397, feb 2008.