University of Pittsburgh | School of Public Health

PeerJ Hubs

Open Access Publishing and Growth Tools for Scholarly Associations

PeerJ

Weighted multiple testing procedures in genome-wide association studies

Ludivine Obry and Cyril Dalmasso

Université Paris-Saclay, CNRS, Univ Evry, Laboratoire de Mathématiques et Modélisation d'Evry, Evry-Courcouronnes, France

ABSTRACT

Multiple testing procedures controlling the false discovery rate (FDR) are increasingly used in the context of genome wide association studies (GWAS), and weighted multiple testing procedures that incorporate covariate information are efficient to improve the power to detect associations. In this work, we evaluate some recent weighted multiple testing procedures in the specific context of GWAS through a simulation study. We also present a new efficient procedure called wBHa that prioritizes the detection of genetic variants with low minor allele frequencies while maximizing the overall detection power. The results indicate good performance of our procedure compared to other weighted multiple testing procedures. In particular, in all simulated settings, wBHa tends to outperform other procedures in detecting rare variants while maintaining good overall power. The use of the different procedures is illustrated with a real dataset.

Subjects Bioinformatics, Genomics, Statistics, Data Science

Keywords False discovery rate, Genome wide association studies, Weighted MTP

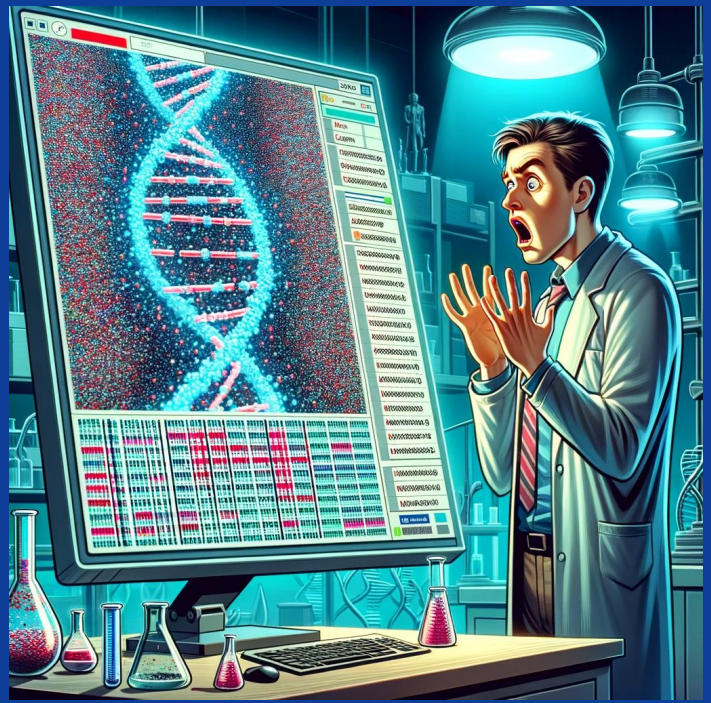
Published in PeerJ June 15, 2023

Presentation by
Dylan Maher
HUGEN 2028
February 12, 2024

University of Pittsburgh | School of Public Health

Motivation

Reminder: Type I Error–false positive
Type II Error–false negative



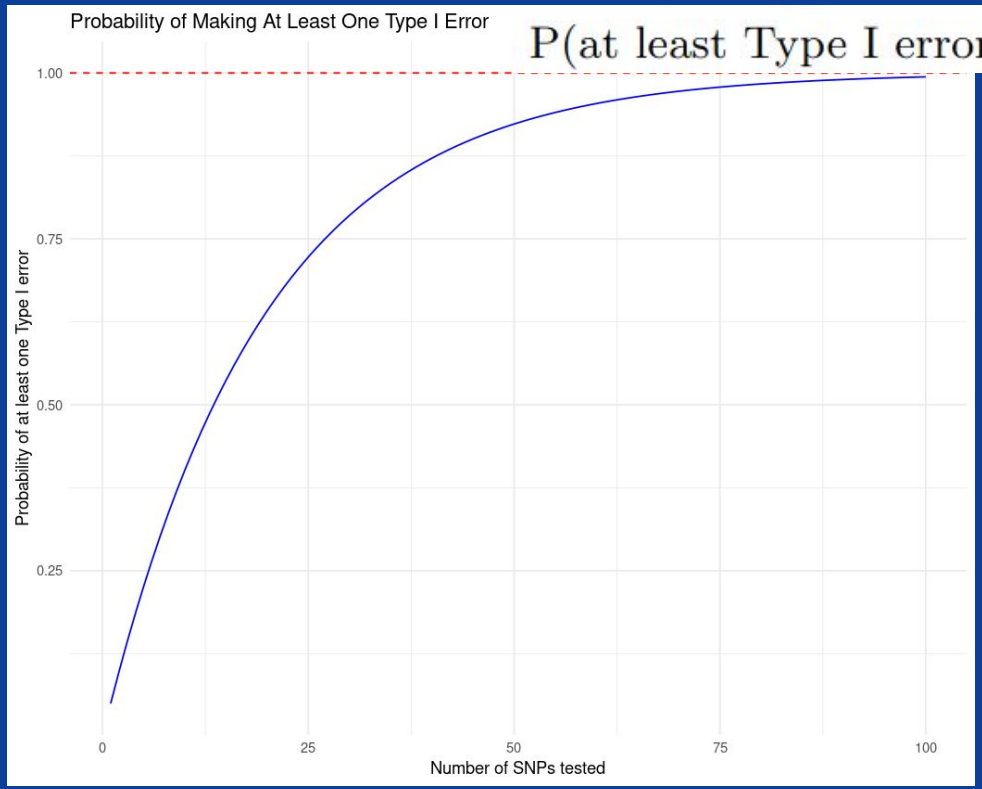
$$P(\text{Type I error}) = \alpha$$
$$P(\text{no Type I error}) = 1 - \alpha$$
$$P(\text{no Type I error in 1 test}) = (1 - \alpha)^1$$
$$P(\text{no Type I error in } n \text{ tests}) = (1 - \alpha)^n$$
$$P(\text{at least Type I error in } n \text{ tests}) = 1 - (1 - \alpha)^n$$

Marginal GWAS involves *n* hypothesis tests (*n*=# SNPs)

SNPs on the order of millions to tens of millions

University of Pittsburgh | School of Public Health

Motivation



Probability of making a Type I Error asymptotes at 1 (rather quickly!)

At 13 SNPs tested, the probability of at least one Type I Error is ~0.5

At 100 SNPs tested, it's ~0.995!

University of Pittsburgh | School of Public Health

Motivation

Two Main Classes of Solutions

Control FWER

Control FDR

Table 1 **Outcomes for m tested hypotheses in a multiple testing situation.**

	H_0 not rejected	H_0 rejected	Total
True H_0	TN	FP	m_0
False H_0 (True H_1)	FN	TP	m_1
	$W = m - R$	R	m

The **FWER** or **family-wise error rate** is the probability of at least one false positive:

$$FWER = P(FP > 0)$$

The **FDR** or **false-discovery rate** is the (expected) proportion of false positives to all positives (true and false):

$$FDR = E\left(\frac{FP}{TP + FP}\right)$$

al., 2017). While correlations between SNPs can substantially deteriorate the performance of many FDR procedures (*Owen, 2005*; *Qiu, Klebanov & Yakovlev, 2005*; *Sarkar, 2006*; *Efron, 2007*; *Neuvial, 2008*), the classical FDR procedures remain valid under different dependence assumptions (*Benjamini & Yekutieli, 2001*; *Farcomeni, 2007*; *Wu et al., 2009*).

University of Pittsburgh | School of Public Health

Basic Idea

Two Main Classes of Solutions

FWER control:

Make sure this number doesn't get to large

How many SNPs are we going to incorrectly call "significant"?

FDR control:

Out of all the SNPs we call "significant," what proportion will be incorrect?

University of Pittsburgh | School of Public Health

Motivation

Two Main Classes of Solutions

- Control FWER
- Control FDR

Traditionally, FWER (specifically, the Bonferroni method) is standard correction used in GWAS, however, as the author's claim:

> Multiple testing procedures controlling the false discovery rate (FDR) are increasingly used in the context of genome wide association studies (GWAS), and weighted multiple

> FDR has become increasingly popular

Is this true?

University of Pittsburgh | School of Public Health

Motivation

$$FWER = P(FP > 0)$$

$$FDR = E\left(\frac{FP}{TP + FP}\right)$$

So, how can ensure sufficient power while maintaining acceptable Type I error control?

FWER is more “stringent”

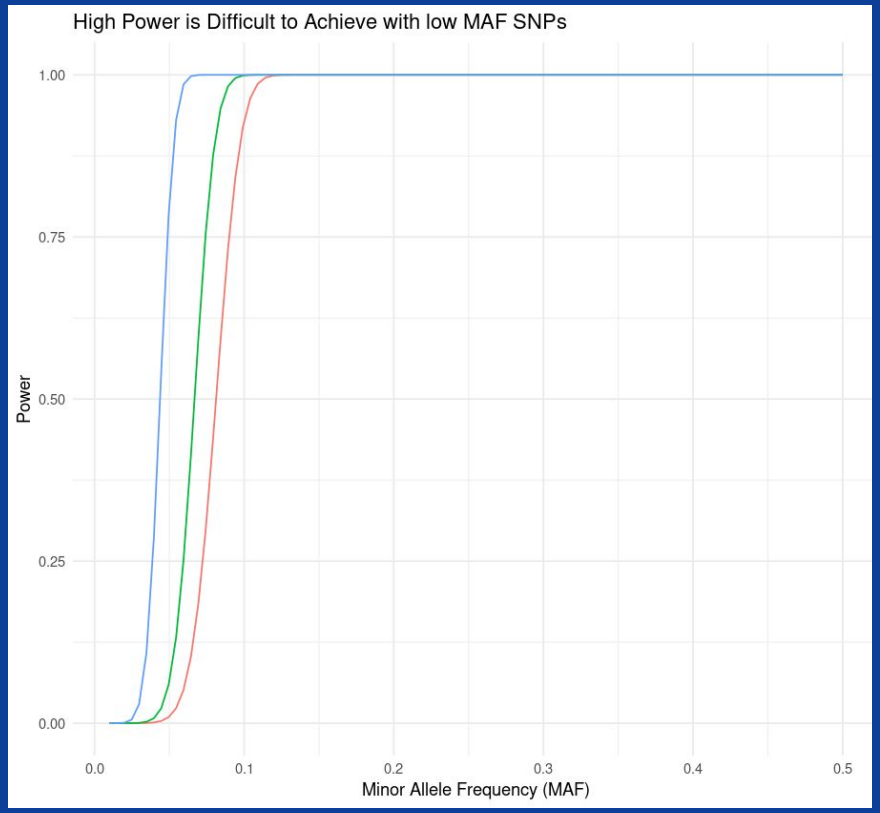
Why is this a concern specifically in the context of GWAS?

2 main reasons–hints

- One is generic statistical concern
- The other is specific to studying genetics (think Bonferroni assumptions)

University of Pittsburgh | School of Public Health

Motivation



Part of issue with power comes down to genetic architecture

In particular, high power is difficult to achieve with rare variants

Burden/collapsing tests aggregate variants and so sacrifice information on individual markers

Alternative solution is weighting strategies

University of Pittsburgh | School of Public Health

Weighting

Idea behind weighting strategies:

Not all “hypotheses” (SNPs) are the same–some are more likely than others

Increase power for SNPs “more likely” to be associated at the expense of others that are “less likely”

The principle of weighted multiple testing procedures is to multiply the thresholds by weights (or equivalently the *p*-values or the test statistics by inverse weights) (*Holm, 1979*; *Benjamini & Hochberg, 1997*; *Genovese, Roeder & Wasserman, 2006*). Thus, the power increases for some individual hypotheses and it decreases for others, while keeping the error criterion control at an average weight equal to 1. In practice, most weighting procedures

University of Pittsburgh | School of Public Health

Weighting

Two Main Classes of Solutions

Control FWER

Control FDR

External Information

wBHa

Adaptive Techniques (information estimated from data)

have recently been introduced (*Ignatiadis et al., 2016*; *Zhang & Chen, 2020*). In a GWAS context, using the MAF as an informative covariate can help to detect rare variants.

What might be advantages/ disadvantages to each?

University of Pittsburgh | School of Public Health

Method

Algorithm 1: *a* Optimization Algorithm

Input: A m-tuple of p-values $P = (p_1, \ldots, p_m)$ and covariates $X = (x_i, \ldots, x_m)$, a nominal level $\alpha \in (0,1)$ for the FDR and a number of folds $K = 100$.

Output: Optimal a

for $k_i = 1, \ldots, K$ **do**

  Sampling with remplacement of $\frac{m}{K}$ hypotheses;

  for $a = 0, 0.1, 0.2, \ldots, 10$ **do**

    Application of wBH procedure at level α with $w(x_i, a) = \frac{m}{\sum_{j=1}^{m} \frac{1}{x_j^a}} \times \frac{1}{x_i^a}$;

    Computation and saving of the numbers of rejections R;

  end

  Saving the values a leading to the maximum of R in an ordered L-tuple ($L \geq 1$) $A = (a_1, \ldots, a_L)$;

  if $L > 1$ **then**

    Computation of the successive differences in A ;

    Definition of interval bounds from differences larger to the step 0.1 ;

    Clustering of the L values of A within the v intervals thus defined ;

    if $v = 1$ **then**

      Saving the maximum value of the vector A ;

    else

      Computation of the length of each interval ;

      if *one of the intervals is longer than the others* **then**

        Saving the maximum value in the longest interval ;

      else

        Saving the maximum value in the interval closest to 1 ;

      end

    end

  end

end

Optimal a obtained by calculating the average of the K values ;

The authors propose a method called wBHa: **w**eighted **B**enjamini-**H**ochberg (**a**daptive)

wBHa procedure

$$w(x_i, a) = \frac{m}{\sum_{j=1}^{m} \frac{1}{x_j^a}} \times \frac{1}{x_i^a}$$

University of Pittsburgh | School of Public Health

Old-School BH

J. R. Statist. Soc. B (1995)
57, *No.* 1, *pp.* 289–300

Controlling the False Discovery Rate: a Practical and Powerful Approach to Multiple Testing

By YOAV BENJAMINI† and YOSEF HOCHBERG

Tel Aviv University, Israel

[Received January 1993. Revised March 1994]

Original method proposed by Benjamini and Hochberg

Basic idea:

1. Rank order p-values (smallest to largest)
2. Multiply them by

$$\frac{i}{m}$$

i ← rank

m ← # of p-values

3. Call new value $p*$
4. Reject null if

$$p* < \alpha$$

University of Pittsburgh | School of Public Health

False discovery control with *p*-value weighting

By CHRISTOPHER R. GENOVESE, KATHRYN ROEDER and LARRY WASSERMAN

Department of Statistics, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213, U.S.A.

Method was extended in 2005 by Genovese, Roeder, and Wasserman (right down the street!)

wBH

Basic idea:

1. Assign each null hypothesis $H_{0,i}$ a non-negative weight

$$w_i : \sum_{i=1}^{m} w_i = m$$

2. Apply BH but replace each p-value with new p-value:

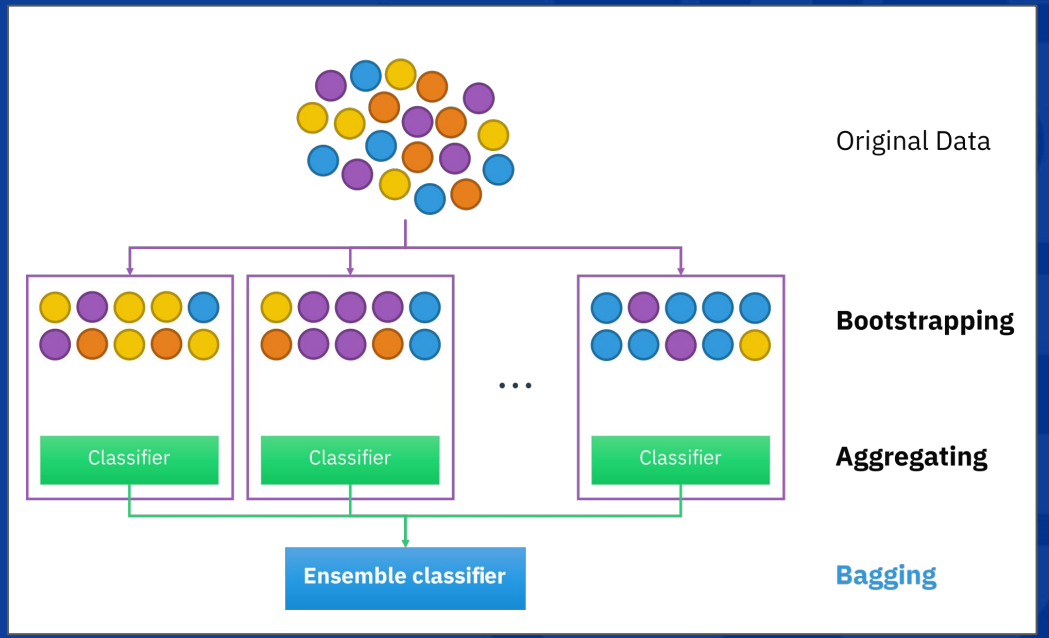
$$p_i = \frac{p_i}{w_i}$$

University of Pittsburgh | School of Public Health

wBHa

Sidenote: "**bagging**" (**b**oostrap **agg**regat**ing**)

1. Create bootstrap samples (sample w/ replacement)
2. Use samples to predict outcome
3. Aggregate outcomes



University of Pittsburgh | School of Public Health

wBHa

wBHa procedure

$$w(x_i, a) = \frac{m}{\sum_{j=1}^{m} \frac{1}{x_j^a}} \times \frac{1}{x_i^a}$$

Assign more “weight” (importance) to hypotheses testing rare variants

Weight varies as a function of MAF (and a)

Two steps:

1. Determine optimal a (maximizes H0 rejections over grid of values)
2. Apply weighted BH

Use bagging to prevent overfitting, take average a over “bags”

Procedure can be used with any covariate, MAF used as example

University of Pittsburgh | School of Public Health

Simulation Study: Genotypes

Genotype matrix: G_{nxm},
$n = 2000$,
$m \in \{8000, 14000, 200000\}$,
$G_{ij} \in \{0, 1, 2\}$,
$G_i^* \sim N_m(0, \Sigma)$
$\Sigma \sim \text{blkdiag}(\Sigma_1, \Sigma_2, ..., \Sigma_k)$
blocks size 10, equicorrelated at ρ
$\rho \in \{0, 0.1, 0.2, 0.3, 0.35, 0.5, 0.75\}$

MAF (causal) divided into four groups with uniforms

- U[0.01, 0.05]
- U[0.05, 0.15]
- U[0.15, 0.25]
- U[0.30, 0.40]

$G_{ij} = 2$ if $q_r < \Phi^{-1}(0, 1)$
$G_{ij} = 1$ if $q_r < \Phi^{-1}(1 - p, 0.1) < q_r$
$G_{ij} = 0$ if $q_r < \Phi^{-1}(p^2, 0.1) < q_r$

The number of causal SNPs (m_1): {5, 10, 15, 20, 25, 50, 100, 150}

Three scenarios for β_j:

- rare causal variants have greater effects
- common variants have greater effects
- all β_j are equal

"Medium-rare" SNPs (hehe)

University of Pittsburgh | School of Public Health

Simulation Study: Phenotypes

Quantitative

$$Y_i = \sum_{j=1}^{m} \beta_j G_{ij} + \varepsilon_i, \quad \varepsilon_i \sim \mathcal{N}(0, \sigma^2)$$

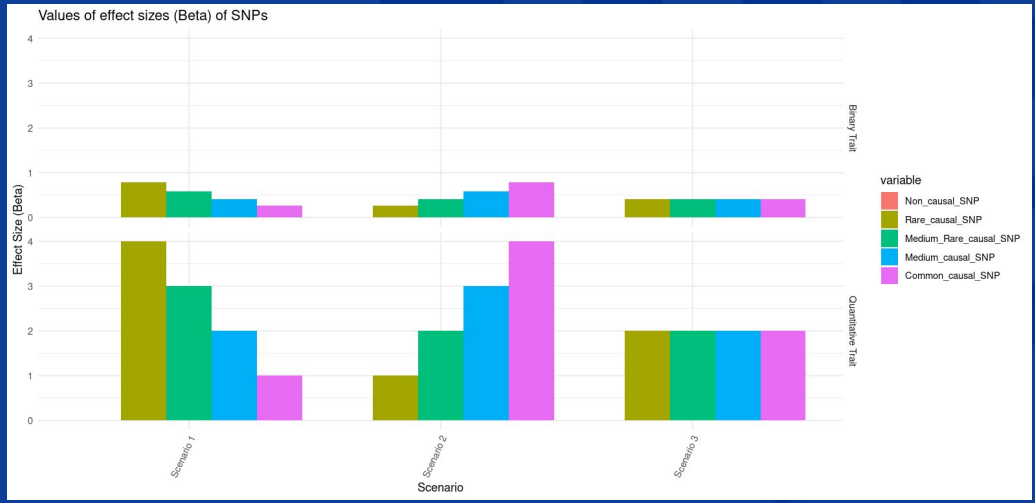
$$\sigma_i^2 = \frac{(R^2 - 1)\sum (G_{ijj} - \bar{Y}_i)^2}{R^2(2 - n)} \quad \text{set} \quad R^2 = 0.2$$

Case-control

$$P(Y_i = 1 | \mathbf{G}_i) = \frac{e^{\beta_0 + \sum_j \beta_j G_{ij}}}{1 + e^{\beta_0 + \sum_j \beta_j G_{ij}}}$$

Table 2 Values of effect sizes (β) of SNPs for quantitative and binary traits into three scenarios.

		Non causal SNP	Rare causal SNP	Medium-Rare causal SNP	Medium causal SNP	Common causal SNP
Quantitative Trait	Scenario 1	0	4	3	2	1
	Scenario 2	0	1	2	3	4
	Scenario 3	0	2	2	2	2
Binary Trait	Scenario 1	0	log(2.2)	log(1.8)	log(1.5)	log(1.3)
	Scenario 2	0	log(1.3)	log(1.5)	log(1.8)	log(2.2)
	Scenario 3	0	log(1.5)	log(1.5)	log(1.5)	log(1.5)



University of Pittsburgh | School of Public Health

Simulation based on Real Dataset

The Search for Host Genetic Factors of HIV/AIDS Pathogenesis in the Post-Genome Era: Progress to Date and New Avenues for Discovery

Bradley E. Aouizerat · C. Leigh Pearce · Christine Miaskowski

307,851 SNPs
605 individuals

MAF bins
[0.01,0.05]: 558
[0.05,0.15]: 4909
[0.15,0.30]: 6674
[0.30,1.00]: 7840

To simulate effect sizes:
Estimated coefficients with wBH
Let effect size = absolute value of quartiles

Three effect size scenarios

1. Rare < Common
2. Common < Rare
3. All equal

University of Pittsburgh | School of Public Health

Simulation Evaluation

Table 3 Procedures compared.

Procedure	R package	Function	Version	Reference
BH	stats	p.adjust	4.2.1	*Benjamini & Hochberg (1995)*
qvalue	qvalue	qvalue	2.28.0	*Storey & Tibshirani (2003)*
FDRreg	FDRreg	FDRreg	0.2.1	*Scott et al. (2015)*
swfdr	swfdr	lm_qvalue	1.22.0	*Boca & Leek (2018)*
IHW	ihw	ihw	1.24.0	*Ignatiadis et al. (2016)*
CAMT	CAMT	camt.fdr	1.1	*Zhang & Chen (2020)*

Table 1 Outcomes for m tested hypotheses in a multiple testing situation.

	H_0 not rejected	H_0 rejected	Total
True H_0	TN	FP	m_0
False H_0 (True H_1)	FN	TP	m_1
	$W = m - R$	R	m

Average power (over 500 iterations)

$$E\left(\frac{TP}{m_1}\right)$$

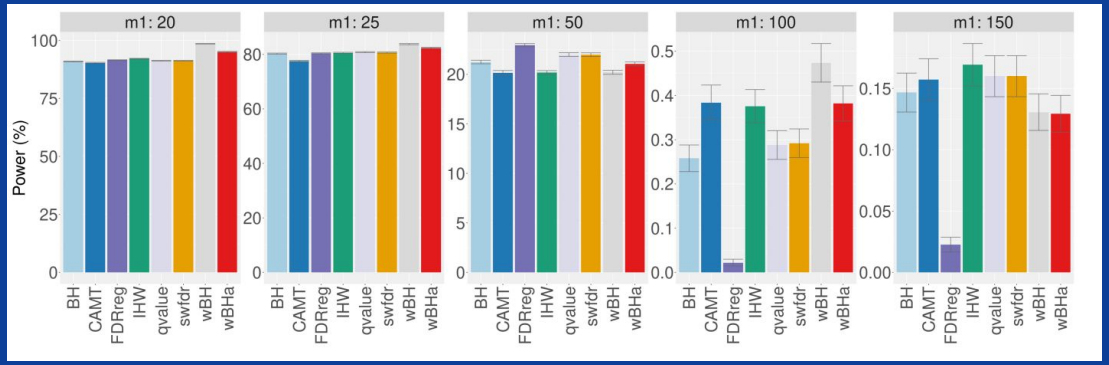
Average power (by subgroup)

$$E\left(\frac{TP_g}{m_{1_g}}\right), g = 1, 2, 3, 4$$

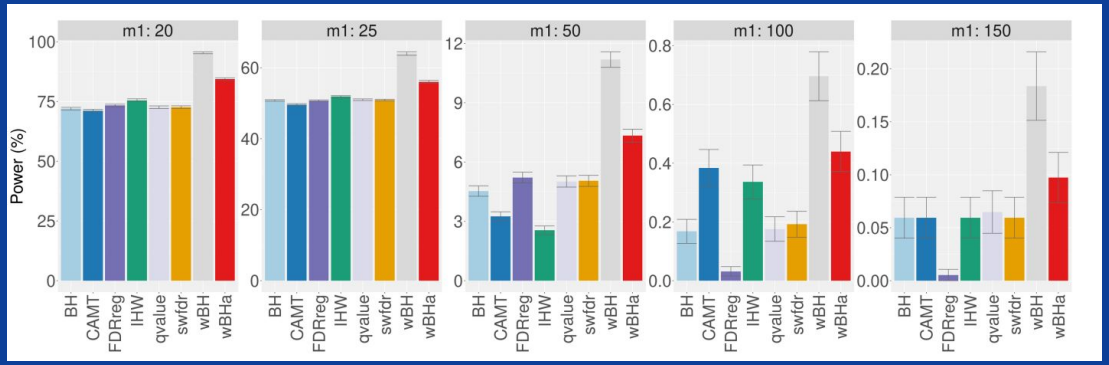
University of Pittsburgh | School of Public Health

Simulation Results

m1 = # of causal SNPs



Overall power comparison in scenario 1, with simulations based on real data, for different m1 values.

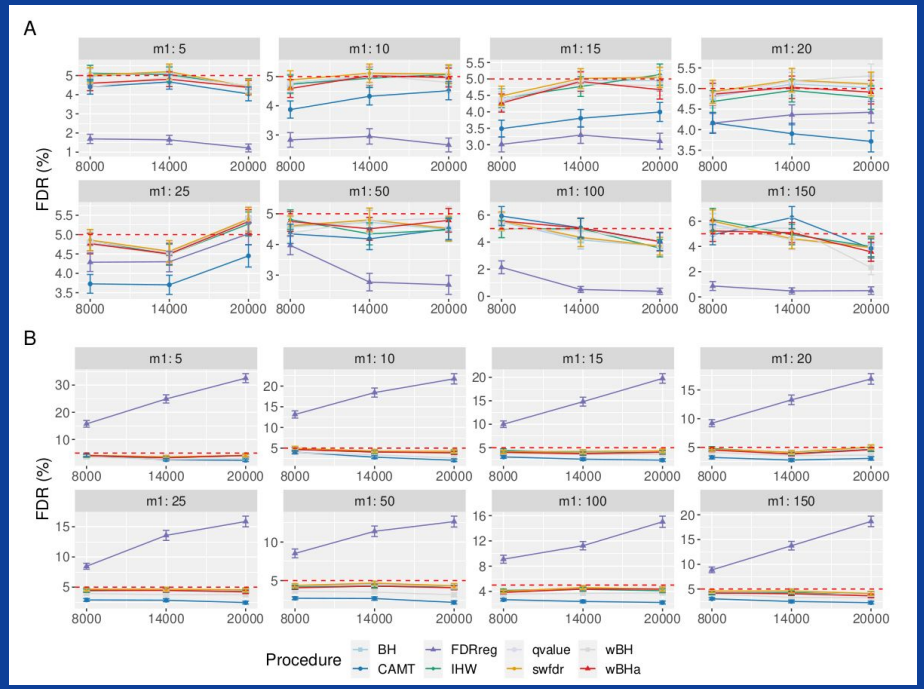


Power comparison in subgroup of rare variants in scenario 1, with simulations based on real data, for different m1 values.

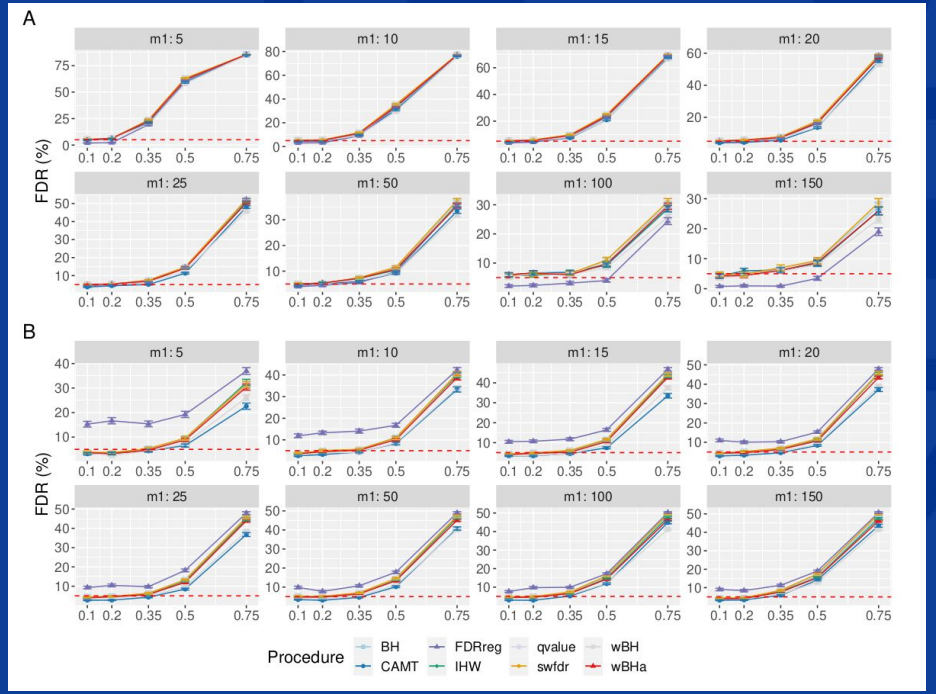
University of Pittsburgh | School of Public Health

Simulation Results

FDR (independent markers)



FDR (correlated markers)



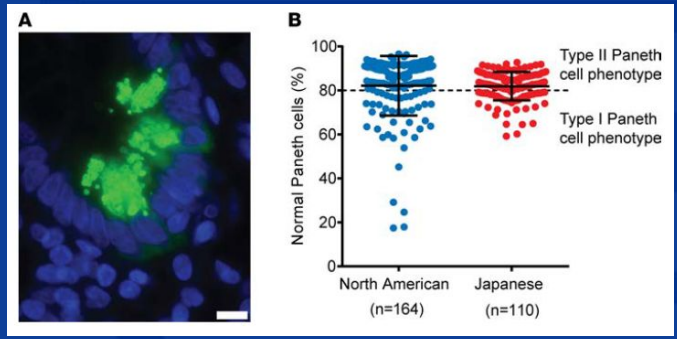
University of Pittsburgh | School of Public Health

Real Dataset

Gene Expression Omnibus (GEO) database

LRRK2 but not _ATG16L1_ is associated with Paneth cell defect in Japanese Crohn's disease patients

Ta-Chiang Liu,[1] Takeo Naito,[2] Zhenqiu Liu,[3] Kelli L. VanDussen,[1] Talin Haritunians,[3] Dalin Li,[3] Katsuya Endo,[2] Yosuke Kawai,[4] Masao Nagasaki,[4] Yoshitaka Kinouchi,[5] Dermot P.B. McGovern,[3] Tooru Shimosegawa,[2] Yoichi Kakuta,[2] and Thaddeus S. Stappenbeck[1]

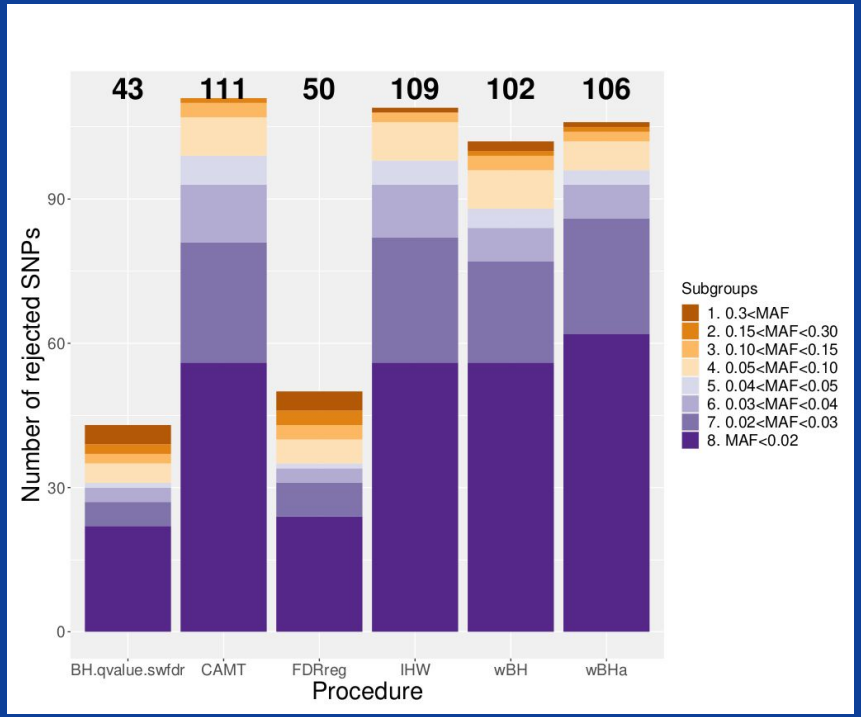


659,636 SNPs (607,720 after QC)
98 individuals
Paneth cell phenotype (small intestine)
Crohn's Disease

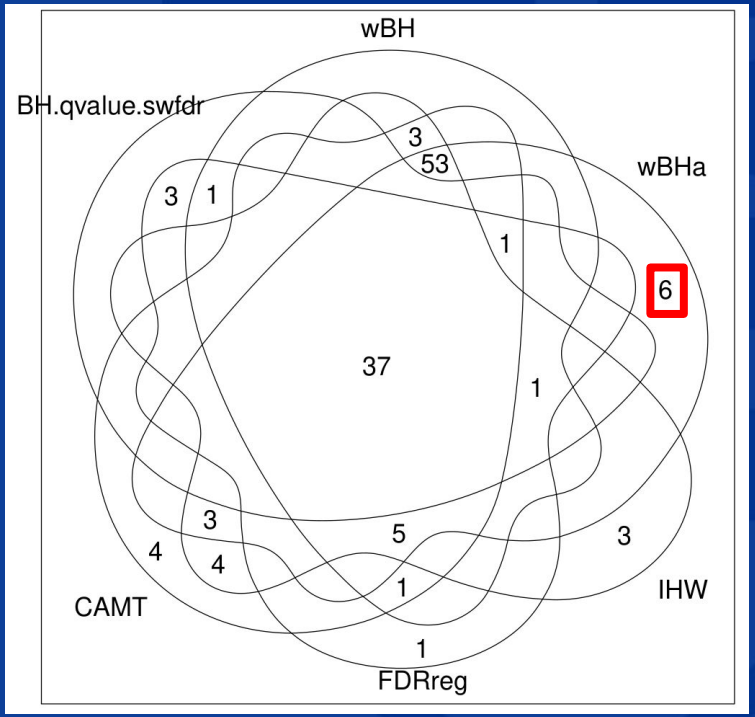
Note that while wBHa is not the most powerful procedure, it identified six specific SNPs (Fig. 8) that could not be selected by the other procedures. Interestingly, two of these SNPs, rs3772479 and rs2270569, are located in the FHIT and KIF9 genes, respectively, which have been reported to play an important role in inflammatory bowel disease (IBD) (Crohn's disease being a type of IBD) (*Skopelitou et al., 2003*; *Xu & Qiao, 2006*; *Wierzbicki et al., 2009*; *Wang et al., 2018*).

University of Pittsburgh | School of Public Health

Real Dataset Results



Rejected SNPs by MAF



Selected SNPs

University of Pittsburgh | School of Public Health

Discussion

Some main takeaways

- wBHa performed well in simulations, good overall power
- IHW & CAMT perform better as proportion of non-null hypotheses increases
- wBH most powerful in all scenarios with rare variants
- In real data analysis, wBHa identified six SNPs undiscoverable by other methods

"Our new procedure wBHa, which showed good performance in all settings, appears to be a good choice for prioritizing rare variants without loss of power."

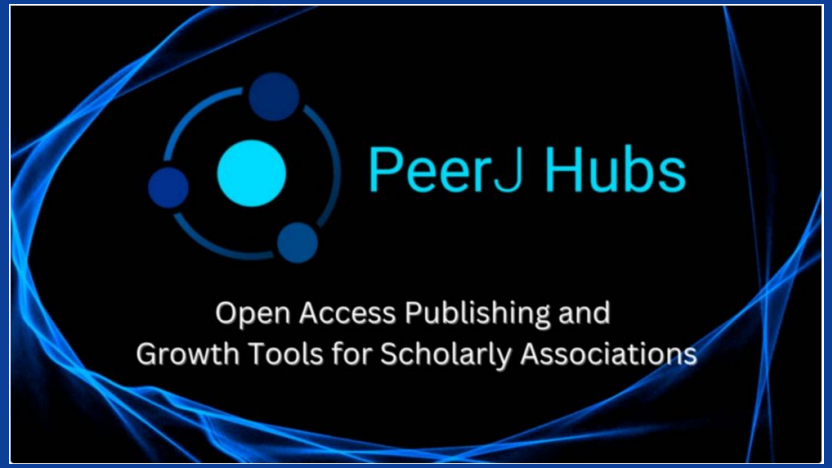
Do we agree?

Authors:

In conclusion, adaptive weighted multiple testing procedures based on informative covariates show great promise in the context of genome-wide association studies. Our new

University of Pittsburgh | School of Public Health

Further Info



AVAILABILITY

The wBHa procedure is implemented in the R package wBHa which is available at https://github.com/obryludivine/wBHa. A second GitHub repository is also available at https://github.com/obryludivine/wBHa_simulation. It contains the programs used to create the simulated datasets and allows our results to be reproduced. These projects have been archived on Zenodo on https://zenodo.org/badge/latestdoi/409590338 and https://zenodo.org/badge/latestdoi/402729574 respectively.

zenodo

R package
install.packages("wBHa")

GitHub repositories available

Zenodo projects archived

Supplemental Information & Author/Reviewer correspondence available on PeerJ

University of Pittsburgh® | School of Public Health

FIN.