Lecture 9: Omnibus Tests, Weighting, Design Considerations

Timothy Thornton and Michael Wu

Summer Institute in Statistical Genetics 2015

Lecture Overview

SKAT vs. Collapsing

- Collapsing tests are more powerful when a large % of variants are causal and effects are in the same direction.
- SKAT is more powerful when a small % of variants are causal, or the effects have mixed directions.
- Both scenarios can happen when scanning the genome.
- Best test to use depends on the underlying biology.

  $\rightarrow$ Difficult to choose which test to use in practice.

We want to develop a unified test that works well in both situations.$\rightarrow$ Omnibus tests

Variable threshold (VT) test

- Previous methods use a fixed threshold for rare variants: $\leq 0.5\%$, $\leq 1\%$, ... $\leq 5\%$?
- Choosing an appropriate threshold can have a huge impact on power

Variable threshold (VT) test

Price AL, Kryukov GV, *et al.*(2010) *AJHG*

- Find the optimal threshold to increase the power.
  - Weight:

$$w_j(t) = \begin{cases} 1 & \text{if} \quad maf_j \leq t \\ 0 & \text{if} \quad maf_j > t \end{cases}$$

  - $C_i(t) = \sum w_j(t) g_{ij}$
  - Test statistics:

$$Z_{max} = max_t Z(t)$$

    where $Z(t)$ is a Z-score of C_i.

P-value Calculations of Variable threshold (VT) test

- Price *et al.*proposed to use permutation to get a p-value
- Lin and Tang (2011) showed that the p-values can be calculated through numerical integration using normal approximation

Variable threshold (VT) test

- More robust than using a fixed threshold.
- Provide information on the MAF ranges of the causal variants.
- Lose power if there exist variants with opposite association directions.

Unified Burden-VC Test

- Burden tests are more powerful when a large % of variants are causal, and all causal variants are harmful (or protective).
- SKAT is more powerful when a small % of variants are causal, or there exist mixed effects.
- Both scenarios can happen across the genome and the underling biology is unknown in advance.

Combine p-values of Burden and SKAT

Derkach A *et al.*(2013) *Genetic Epi*, 37:110-121

- Fisher method:

$$Q_{Fisher} = -2\log(P_{Burden}) - 2\log(P_{SKAT})$$

- Q_{Fisher} follows χ^2 with 4 d.f when these two p-values are independent
- Since they are not independent, p-values are calculated using resampling
- Mist (Sun et al. 2013) modified the SKAT test statistics to make them independent

Combine Test Statistics: Unified Test Statistics

Lee *et al.*(2012). *Biostatistics*

- Combined Test of Burden tests and SKAT

$$Q_\rho = (1-\rho)Q_{SKAT} + \rho Q_{Burden}, \quad 0 \leq \rho \leq 1.$$

- Q_ρ includes SKAT and burden tests.
  - $\rho = 0$: SKAT
  - $\rho = 1$: Burden

Derivation of the Unified Test Statistics

- Model:

$$g(\mu_i) = \mathbf{X}_i\boldsymbol{\alpha} + \mathbf{G}_i\boldsymbol{\beta}$$

  where β_j/w_j follows any arbitrary distribution with mean 0 and variance τ and the correlation among β_j's is ρ.
- Special cases:
  - SKAT: $\rho = 0$
  - Burden: $\rho = 1$
  - Combined: $0 \leq \rho \leq 1$

Derivation of the Unified Test Statsitics

- Q_ρ is a test statistic of the SKAT with $corr(\beta) = \mathbf{R}(\rho)$:
  - $\mathbf{R}(\rho) = (1-\rho)\mathbf{I} + \rho \underline{1}\underline{1}'$ (compound symmetric)
  - $\mathbf{K}_\rho = \mathbf{GWR}(\rho)\mathbf{WG}'$.

$$\begin{aligned} Q_\rho &= (\mathbf{y} - \hat{\boldsymbol{\mu}})' \mathbf{K}_\rho (\mathbf{y} - \hat{\boldsymbol{\mu}}) \\ &= (1-\rho) Q_{SKAT} + \rho Q_{Burden} \end{aligned}$$

Adaptive Test (SKAT-O)

- Use the smallest p-value from different ρs:

$$T = \inf_{0 \leq \rho \leq 1} P_\rho.$$

  where P_ρ is the p-value of Q_ρ for given ρ.

- Test statistic:

$$T = minP_{\rho_b}, \quad 0 = \rho_1 < \ldots < \rho_B = 1.$$

Adaptive Test (SKAT-O)

- Q_ρ is a mixture of two quadratic forms.

$$\begin{aligned} Q_\rho &= (1-\rho)(\mathbf{y}-\hat{\boldsymbol{\mu}})' GWWG'(\mathbf{y}-\hat{\boldsymbol{\mu}})' \\ &\quad + \rho(\mathbf{y}-\hat{\boldsymbol{\mu}})' GW\underline{1}\underline{1}'WG'(\mathbf{y}-\hat{\boldsymbol{\mu}}) \\ &= (1-\rho)(\mathbf{y}-\hat{\boldsymbol{\mu}})' K_1(\mathbf{y}-\hat{\boldsymbol{\mu}})' + \rho(\mathbf{y}-\hat{\boldsymbol{\mu}})' K_2(\mathbf{y}-\hat{\boldsymbol{\mu}}) \end{aligned}$$

- Q_ρ is asymptotically equivalent to

$$(1-\rho)\kappa + a(\rho)\eta_0,$$

where and $\eta_0 \sim \chi^2_1$, κ approximately follows a mixture of χ^2.

SKAT-O

- Q_ρ is the asymptotically same as the sum of two independent random variables.

$$(1-\rho)\kappa + a(\rho)\eta_0$$

  - $\eta_0 \sim \chi^2_1$
  - Approximate κ via moments matching.

- P-value of T:

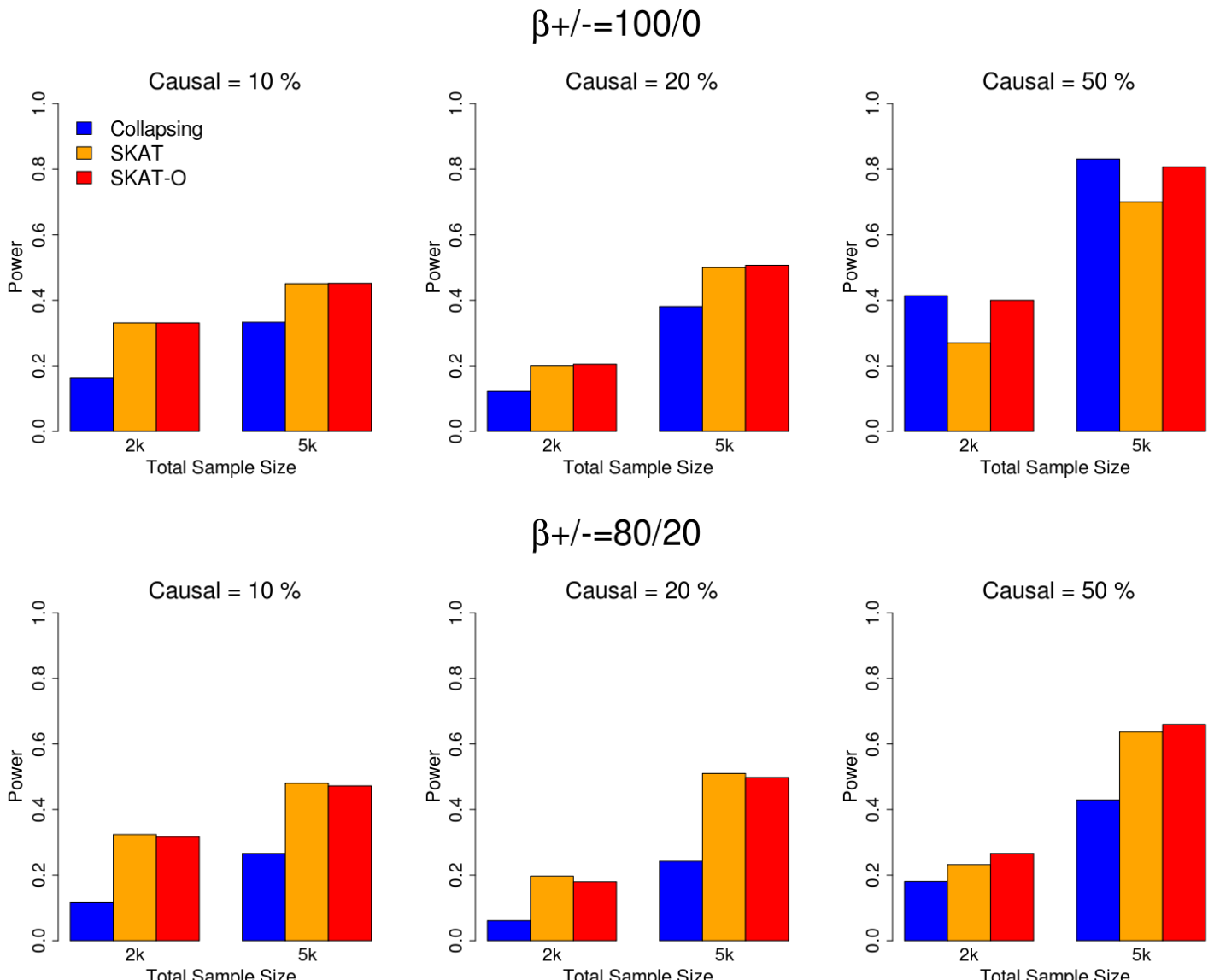
$$\begin{aligned} & 1 - Pr\left\{Q_{\rho_1} < q_{\rho_1}(T), \ldots, Q_{\rho_b} < q_{\rho_b}(T)\right\} \\ & = 1 - E\left[Pr\left\{(1-\rho_1)\kappa + a(\rho_1)\eta_0 < q_{\rho_1}(T), \ldots | \eta_0\right\}\right] \\ & = 1 - E\left[P\left\{\kappa < min\{(q_{\rho_v}(T)) - a(\rho_v)\eta_0)/(1-\rho_v)\} | \eta_0\right\}\right], \end{aligned}$$

where $q_\rho(T) =$ quantile function of Q_ρ

Simulation

- Simulate sequencing data using COSI
- 3kb randomly selected regions.
- Percentages of causal variants = 10%, 20%, or 50%.
- $(\beta_j > 0)$% among causal variants = 100% or 80%.
- Three methods
  - Burden test with beta(1,25) weight
  - SKAT
  - SKAT-O

Simulation


β+/-=100/0
Causal = 10 %
Causal = 20 %
Causal = 50 %
Collapsing
SKAT
SKAT-O
Power
0.0
0.2
0.4
0.6
0.8
1.0
2k
5k
Total Sample Size
β+/-=80/20
Causal = 10 %
Causal = 20 %
Causal = 50 %

Simulation

- SKAT is more powerful than Burden test (Collapsing) when
  - Existence of $+/-\beta$s
  - Small percentage of variants are causal variants
- Burden test is more powerful than SKAT when
  - All βs were positive and a large proportion of variants were casual variants
- SKAT-O is robustly powerful under different scenarios.

Summary

- Region based tests can increase the power of rare variants analysis.
- Relative performance of rare variant tests depends on underlying disease models
- The combined test (omnibus test), e.g, SKAT-O, is robust and powerful in different scenarios

MAF based weighting

- It is generally assumed that rarer variants are more likely to be causal variants with larger effect sizes.
- Simple thresholding is widely used.

$$w(MAF_j) = \begin{cases} 1 & \text{if} \quad MAF_j < c \\ 0 & \text{if} \quad MAF_j \geq c \end{cases}$$

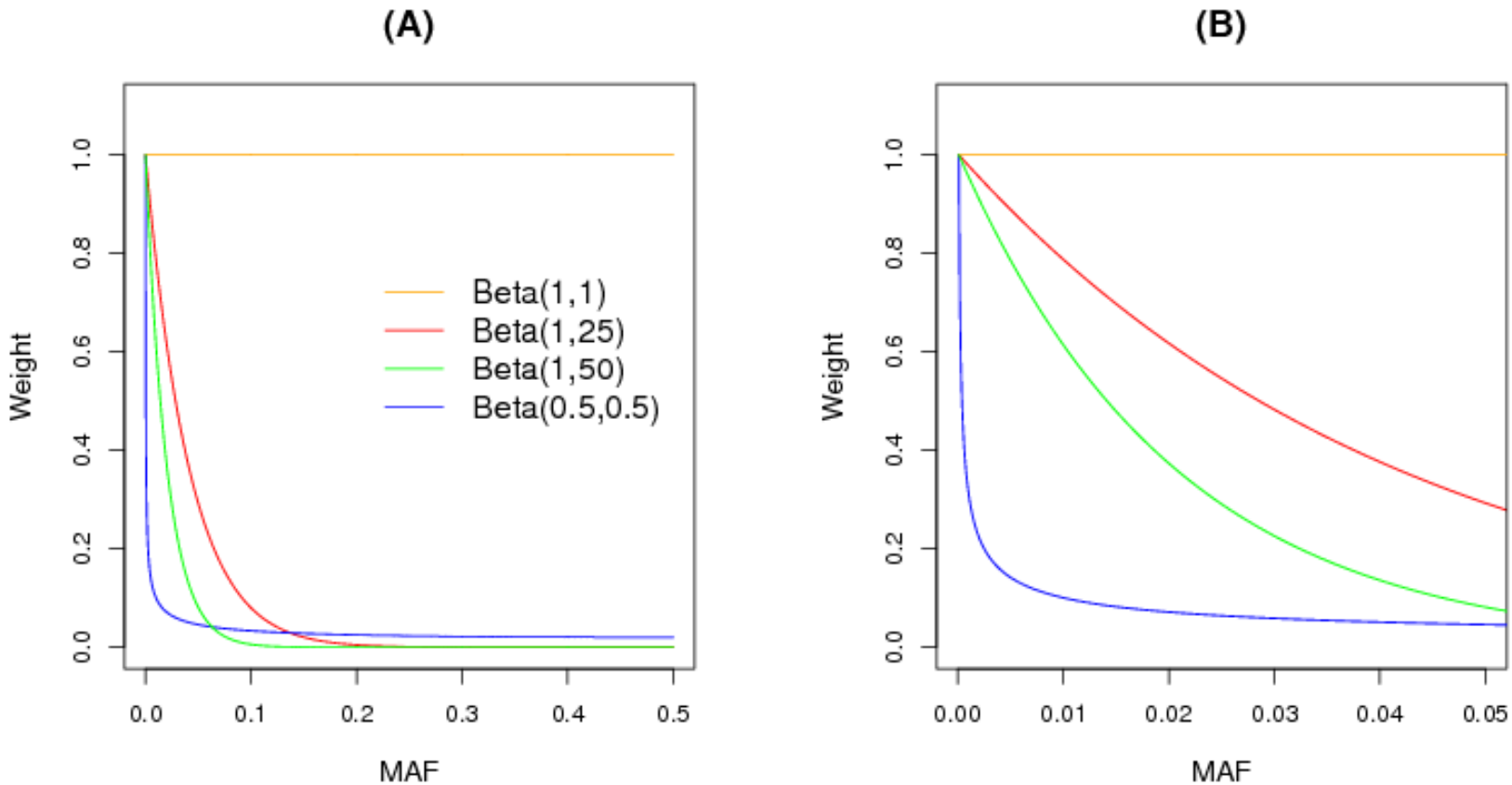
MAF based weighting

- Instead of thresholding, continuous weighting can be used to upweight rarer variants.
- Ex: Flexible beta density function.

$$w(MAF_j) = (MAF_j)^{\alpha-1}(1 - MAF_j)^{\beta-1}$$

  - $(\alpha = 0.5, \beta = 0.5)$: Madsen and Browning weight
  - $(\alpha = 1, \beta = 1)$: Flat weight

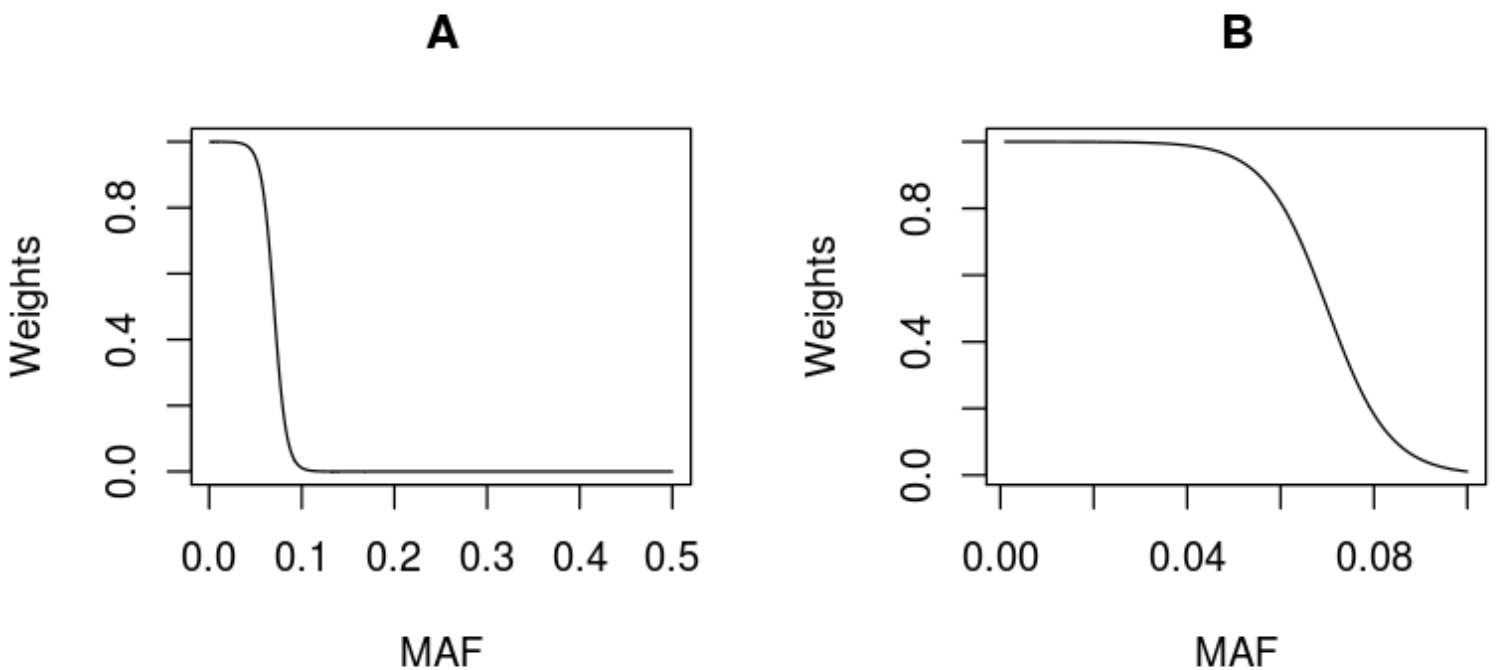
MAF based weighting- beta weight



MAF based weighting- logistic weight

- Soft-thresholding.

$$w(maf_j) = exp((\alpha - maf_j)\beta)/\{1 + exp((\alpha - maf_j)\beta\}$$



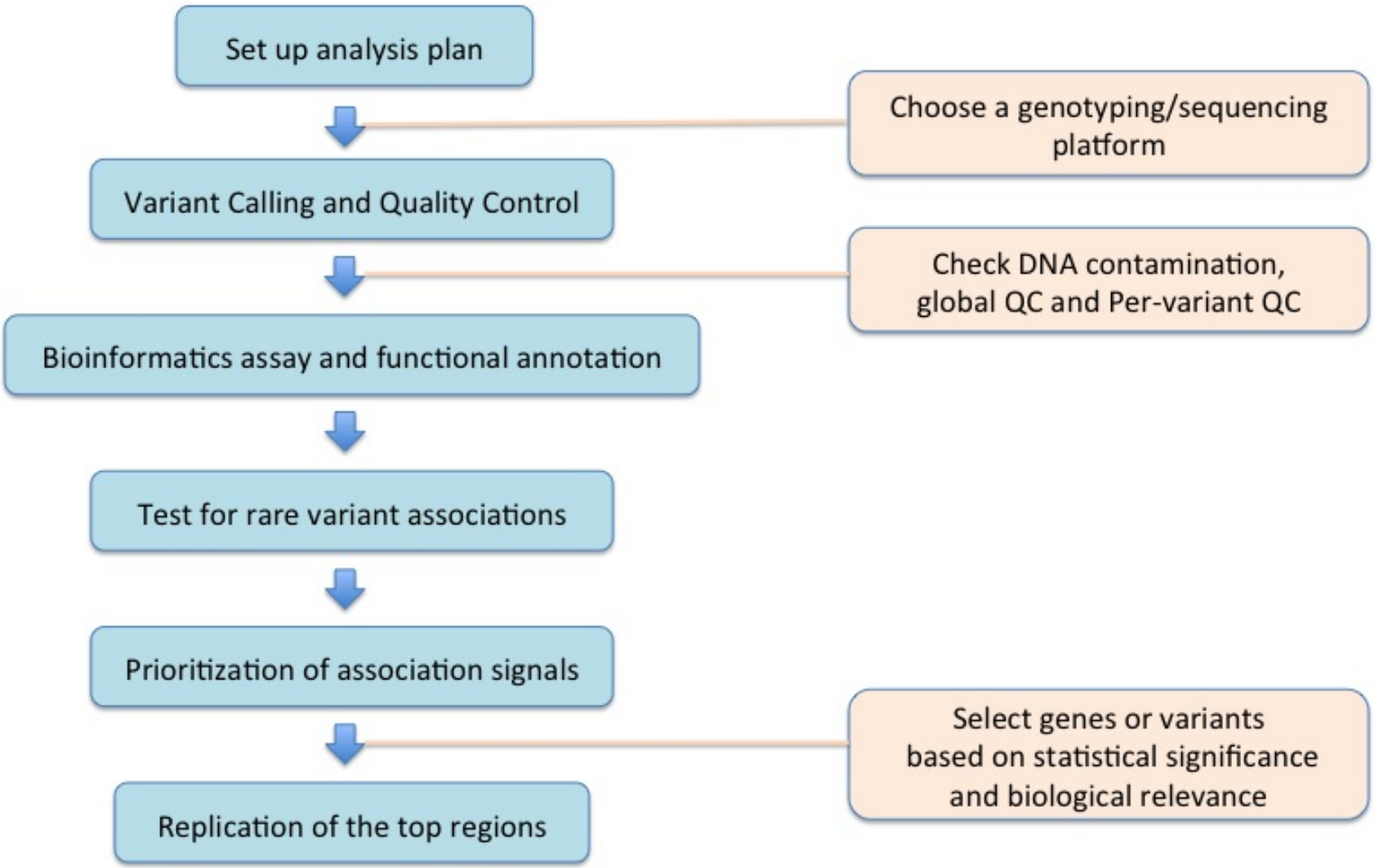
Weighting Using Functional information

- Variants have different functionalities.
  - Non-synonymous mutations (e.g. missense and nonsense mutations) change the amino-acid (AA) sequence.
  - Synonymous mutations do not change AA sequence.

Weighting Using Functional information

- Bioinformatic tools to predict the functionality of mutations.
  - Polyphen2 (http://genetics.bwh.harvard.edu/pph2/)
  - SIFT (http://sift.jcvi.org/)
- Test only functional mutations can increase the power.

Data Processing and Analysis Flowchart



Genotyping Platforms

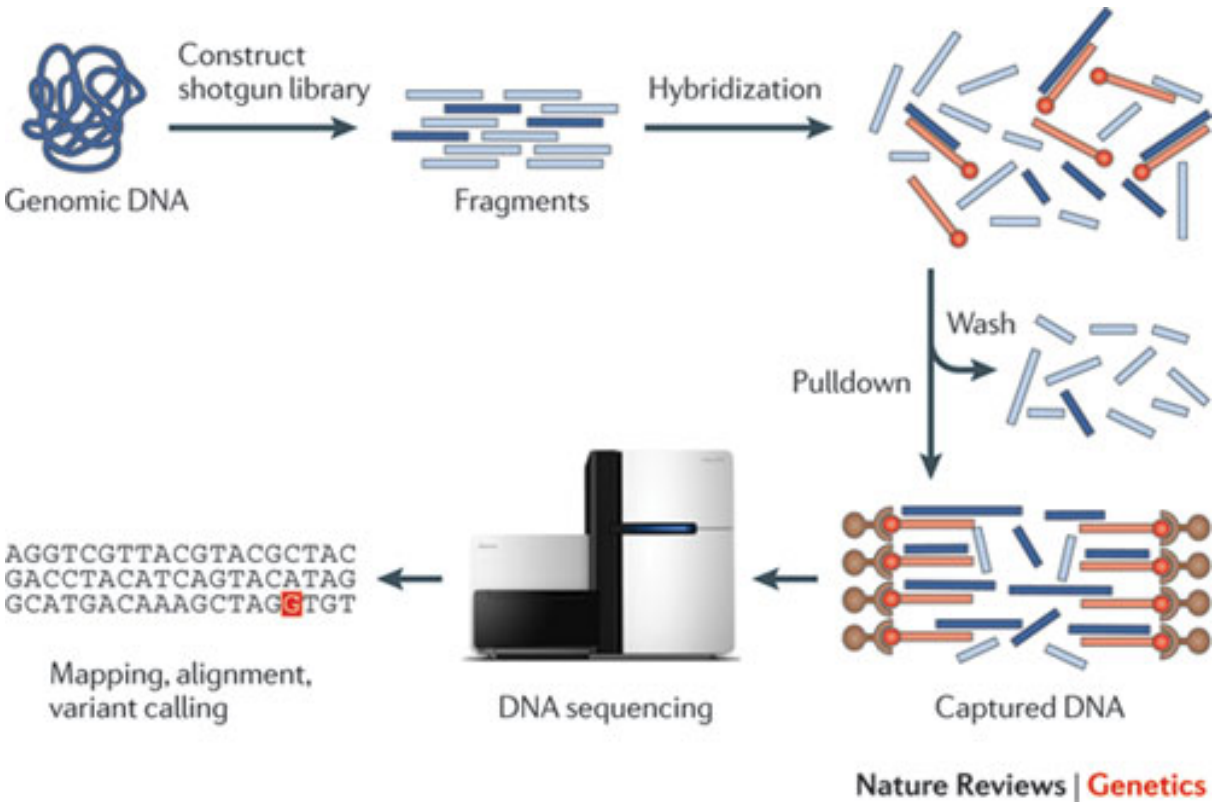
- High depth whole genome sequencing is the most informative, however it is currently expensive.
- Alternative sequencing designs and genotyping platforms
  - Low depth sequencing
  - Exome seqeuncing
  - High coverage microarrays (Exome chip)
  - Imputation

Low depth whole genome sequencing

- Sequencing $7 \sim 8$ samples at low depth (4x) instead of 1 sample at high depth (30x)
- Low depth sequencing
  - Relatively affordable
  - LD based genotyping: leverage information across individuals to improve genotype accuracy.
  - 1000 Genome (4x) and UK 10K (6x) used low depth sequencing.
- Cons:
  - Subject to appreciable sequencing errors

Exome sequencing

- Restrict to the protein coding region ($1 \sim 2\%$ of genome (30 Mbps)).



Bamshad et al. (2011), NRG

Exome sequencing

- Focus on the high value portion of the genome
- Relatively cost effective
- Cons: Only focus on the exome
  - Most of GWAS hits lie in non-exomic regions
  - Many non-coding regions have biological functions

Exome array

- Using variants discovered in 12,000 sequenced exome
- Low cost ($10 \sim 20x$ less than Exome sequencing)
  - 250K non-synonymous variants
  - 12K splicing variants
  - 7K stop altering variants
- Cons:
  - Cannot investigate very rare variants.
  - Limited coverages for non-European populations

GWAS chip + Imputation

- Imputation: Estimate genotypes using reference samples
  - Imputation accuracy increases as the number of reference samples increases
- No additional experiment cost
- Cons:
  - Low accuracy of imputed rare variants

Summary

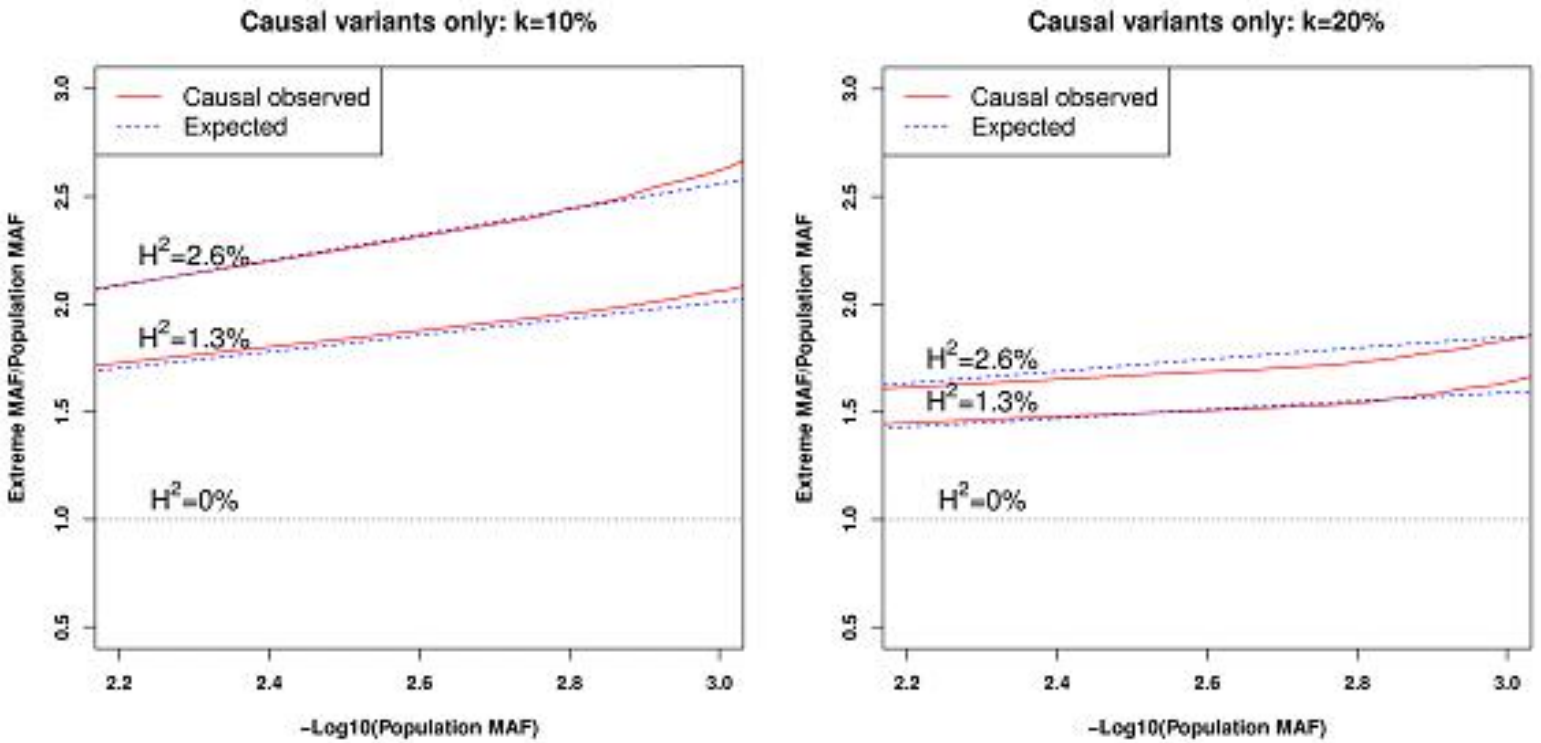
	Advantage	Disadvantage
High-depth WGS	Can identify nearly all variants in genome with high confidence.	Currently very expensive.
Low-depth WGS	Cost-effective, useful approach for association mapping.	Limited accuracy
Whole exome sequencing	Can identify all exomic variants; less expensive than WGS.	Limited to the exome.
GWAS chip + Imputation	Low cost.	Lower accuracy of imputed rare variants.
Exome chip (custom array)	Much cheaper than exome sequencing.	Limited coverage for very rare variants and for non-Europeans. Limited to target regions.

Extreme phenotype sampling

- Rare causal variants can be enriched in extreme phenotypic samples
- Given the fixed budget, increase power by sequencing extreme phenotypic samples.

Enrichment of causal rare variants in phenotypic extremes

- Estimated folds increase of the observed MAFs of causal variants (k% high/low sampling, H^2=Heritability).



Extreme phenotypic sampling

- Continuous traits:
  Select individuals with extreme trait values after adjusting for covariates.
- Binary traits:
  Select individuals on the basis of known risk factors
  - Ex. T2D : family history, early onset, low BMI

Extreme phenotypic sampling

- Extreme continuous phenotype (ECP) can be dichotomized, and then any testing methods for binary traits can be used.
- But dichotomization can cause a loss of information and can decrease the power.
- Methods modeling ECP as truncated normal distribution has been developed (Barnett, et al, 2013, Gen. Epid).

Power/Sample Size calculation

- Power/Sample size calculation is essential to design future sequencing studies.
- Input information:
- Region information
  - LD structure and MAF spectrum.
  - Region size to test.

Power/Sample Size calculation

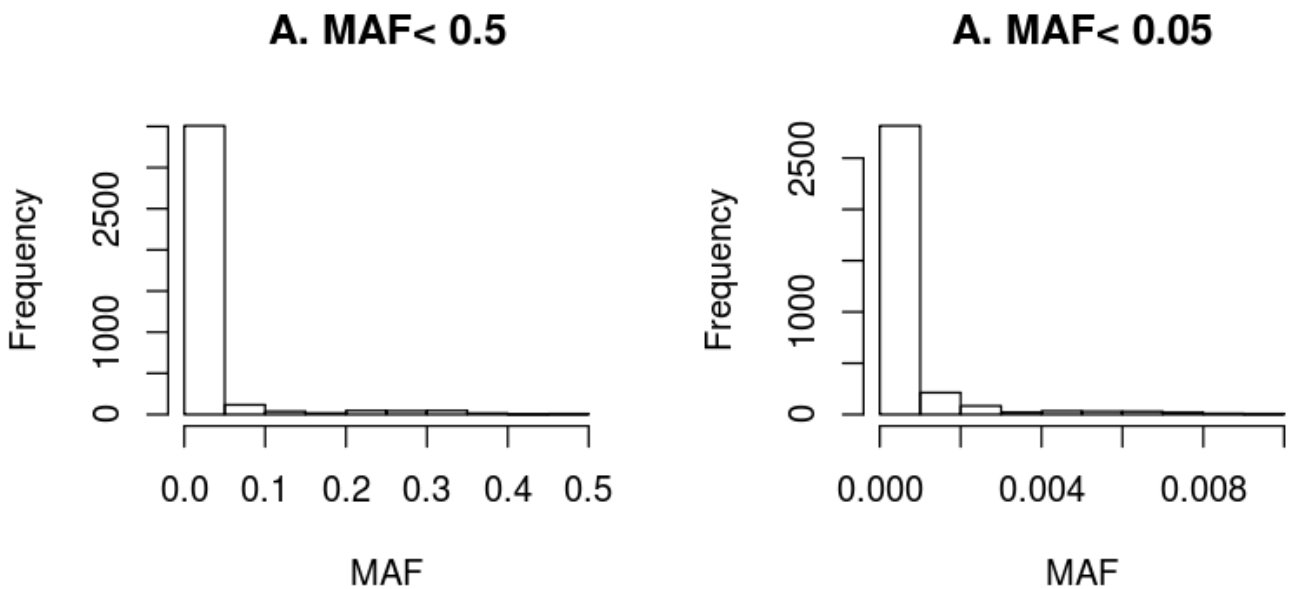
- Causal variant Information
  - Effect size (continuous traits), or Odds ratio (binary traits).
  - % of rare variants be causal.
  - % of causal variants with negative association direction.
- Binary traits
  - Case/Control Ratio.
  - Prevalence

Practical Points: SKAT Power Calculations

- Region information
  - Either simulated haplotypes or sample haplotypes from preliminary data.
  - The SKAT package provides 10,000 haplotypes over a 200 kb region generated by the coalescent simulator (COSI).

MAF spectrum

- MAF spectrum of the simulated haplotypes
- Most of SNPs have very low MAFs.



Practical Points: Power/Sample Size calculations

- Causal Variant Information:
  - To use $\log_{10}$ function $(-c \log_{10}(MAF))$ for the effect sizes or log odds ratio.
  - c is a parameter to determine the strength of association.
    - Ex: $c = 1$
      $\beta = 2$ or $\log(OR) = 2$ for a variant with MAF=0.01
      $\beta = 4$ or $\log(OR) = 4$ for a variant with MAF=10^{-4}.

Practical Points: Power/Sample Size calculations

- In SKAT package, you can set c using the MaxOR (OR for MAF $= 10^{-4}$) or MaxBeta (β for MAF $= 10^{-4}$).

Practical Points: Power/Sample Size calculations

- Power depends on LD structure of the region and MAFs of the causal variants.
- We are interested in estimating power in multiple regions and multiple sets of causal variants selected from a certain disease model.
  - We estimate an average power.
  - Approximately 100 $\sim$ 500 sets of regions/causal variants are needed to estimate the average power stably.