Genetics Studies of Multivariate Traits

Heping Zhang

Department of Epidemiology and Public Health
Yale University School of Medicine
Presented at
Southern Regional Council on Statistics
Summer Research Conference

June 7, 2011

Outline

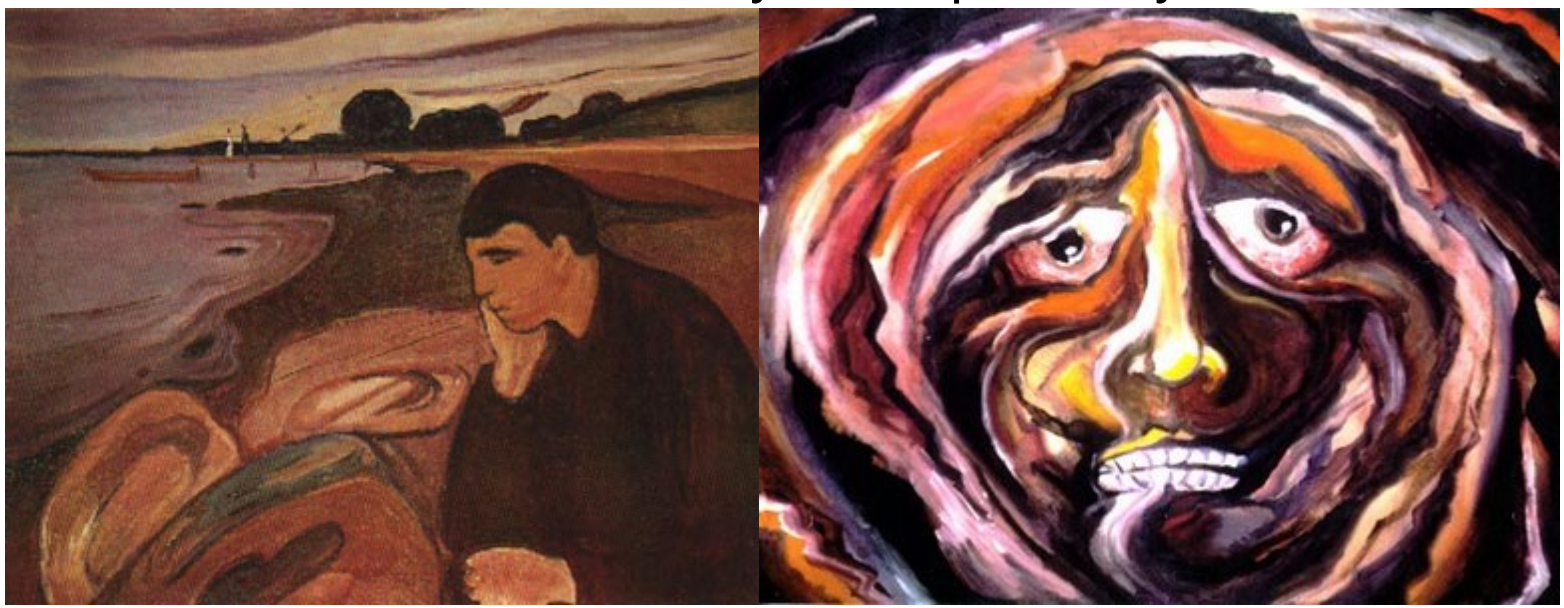
1. Background
   - Comorbidity
   - Disorders, Genes and Covariates
2. Weighted Association Test
   - Generalized Kendall's Tau
   - Asymptotic Distribution and Power
   - Application to WTCCC Bipolar Disorder Data
3. Maximum Weighted Association Test
   - Asymptotic Distribution
   - Simulations
   - Application to COGA Data
4. Conclusions

Outline

1. Background
   - Comorbidity
   - Disorders, Genes and Covariates
2. Weighted Association Test
   - Generalized Kendall's Tau
   - Asymptotic Distribution and Power
   - Application to WTCCC Bipolar Disorder Data
3. Maximum Weighted Association Test
   - Asymptotic Distribution
   - Simulations
   - Application to COGA Data
4. Conclusions

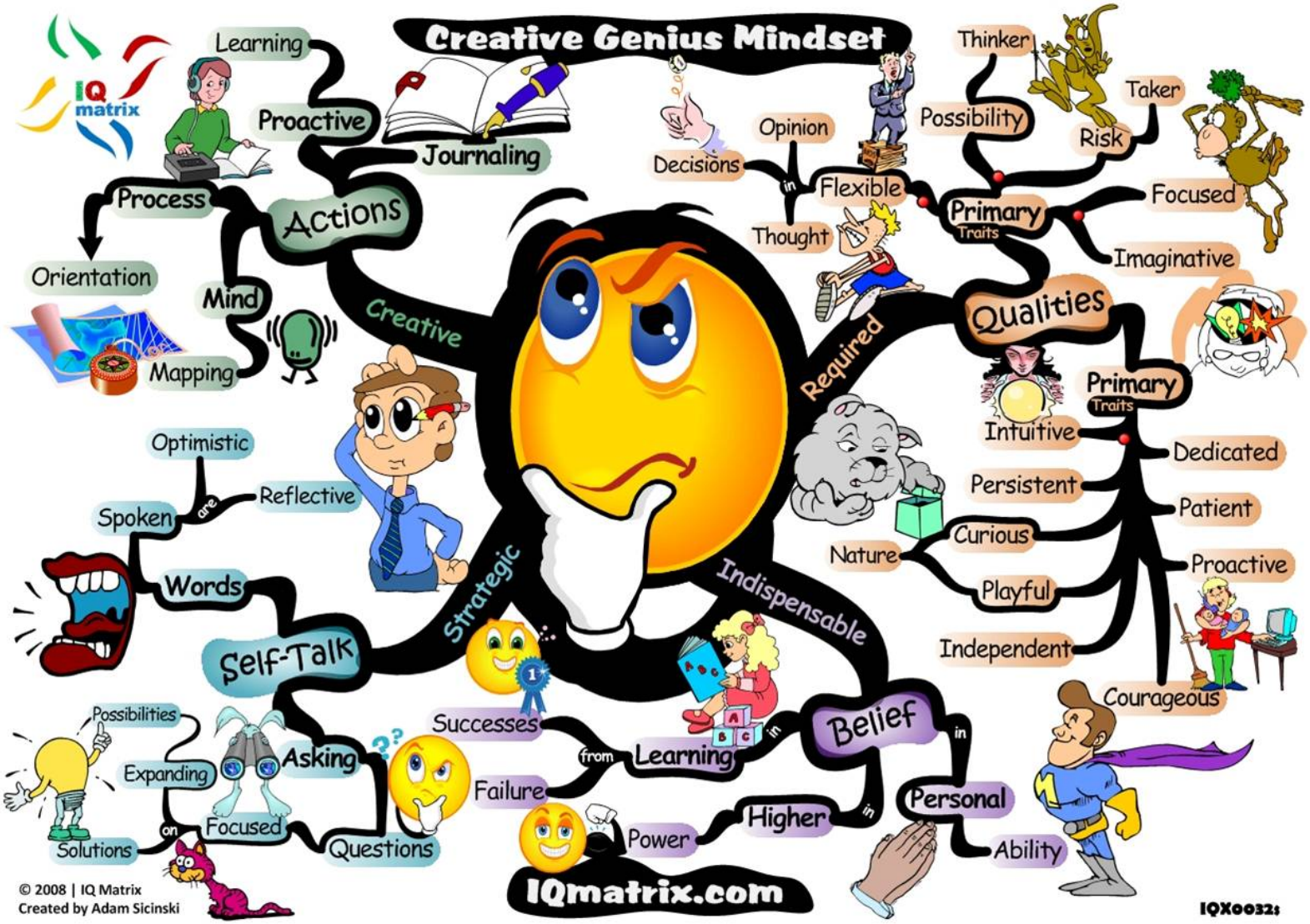
Comorbidity

Multiple disorders or illnesses occur in the same person, simultaneously or sequentially



Source: www.depressioncell.com; www.depressiondodging.com

Comorbidity



Source: aasets.lifehack.org

Comorbidity

Is there a relationship between childhood ADHD and later drug abuse? See page 2.

NIDA NATIONAL INSTITUTE ON DRUG ABUSE

Research Report Series

Comorbidity:
Addiction and Other Mental Illnesses

from the director:

Comorbidity is a topic that our stakeholders—patients, family members, health care professionals, and others—frequently ask about. It is also a topic about which we have insufficient information, so it remains a research priority for NIDA. This Research Report provides information on the state of the science in this area. Although a variety of diseases commonly co-occur with drug abuse and addiction (e.g., HIV, hepatitis C, cancer, cardiovascular disease), this report focuses only on the comorbidity of drug use disorders and other mental illnesses.*

To help explain this comorbidity, we need to first recognize that drug addiction is a mental illness. It is a complex brain disease characterized by compulsive, at times uncontrollable drug craving, seeking, and use despite devastating consequences—behaviors that stem from drug-induced changes in brain structure and function. These changes occur in some of the same brain areas that are disrupted in other mental disorders, such as depression, anxiety, or schizophrenia. It is therefore not surprising that population surveys show a high rate of co-occurrence, or comorbidity, between drug addiction and other mental illnesses. While we cannot always prove a connection or causality, we do know that certain mental disorders are established risk factors for subsequent drug abuse—and vice versa.

It is often difficult to disentangle the overlapping symptoms of drug addiction and other mental illnesses, making diagnosis and treatment complex. Correct diagnosis is critical to ensuring appropriate and effective treatment. Ignorance of or failure to treat a comorbid disorder can jeopardize a patient's chance of recovery. We hope that our enhanced understanding of the common genetic, environmental, and neural bases of these disorders—and the dissemination of this information—will lead to improved treatments for comorbidity and will diminish the social stigma that makes patients reluctant to seek the treatment they need.

Nora D. Volkow, M.D.
Director
National Institute on Drug Abuse

What Is Comorbidity?

When two disorders or illnesses occur in the same person, simultaneously or sequentially, they are described as comorbid. Comorbidity also implies interactions between the illnesses that affect the course and prognosis of both.

continued inside

*Since the focus of this report is on comorbid drug use disorders and other mental illnesses, the terms "mental illness" and "mental disorders" will refer here to disorders other than substance use disorders, such as depression, schizophrenia, anxiety, and mania. The terms "dual diagnosis," "mentally ill chemical abuser," and "co-occurrence" are also used to refer to drug use disorders that are comorbid with other mental illnesses.

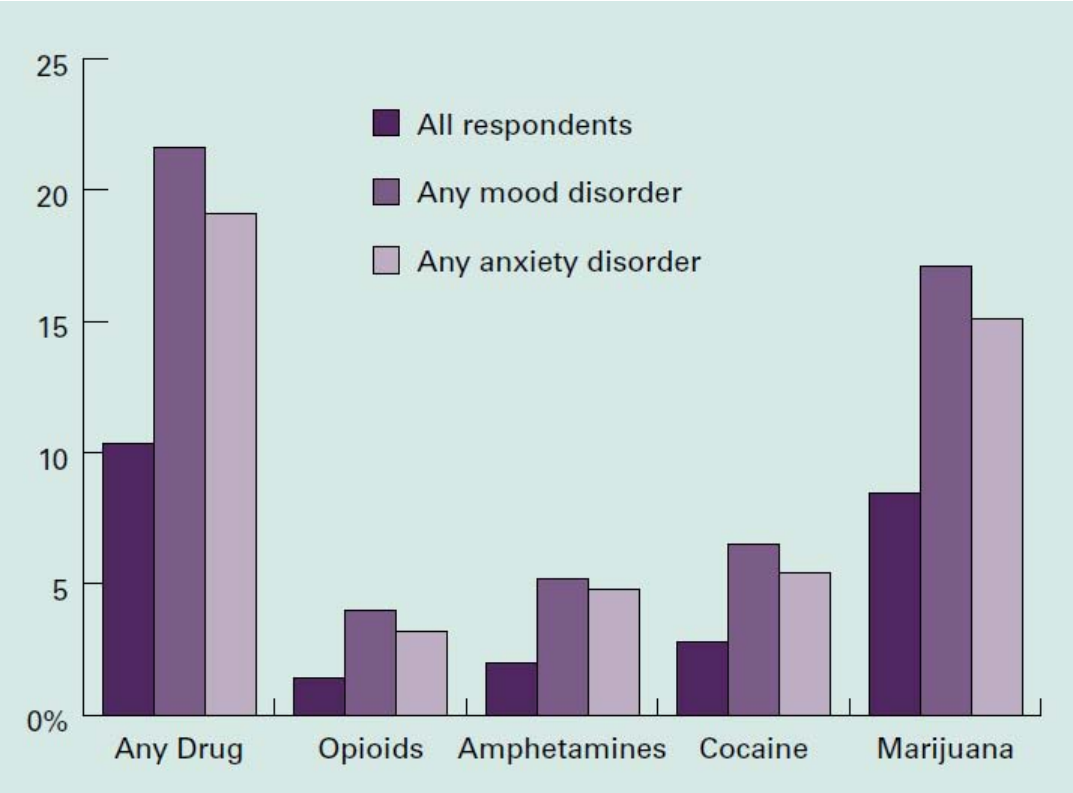
U.S. Department of Health and Human Services | National Institutes of Health

Dr. Volkow, Director, NIDA: Comorbidity is a topic that our stakeholders–patients, family members, health care professionals, and others–frequently ask about. It is also a topic about which we have insufficient information, so it remains a research priority for NIDA.

Source: www.nida.nih.gov

Possible Mechanisms for Comorbidity

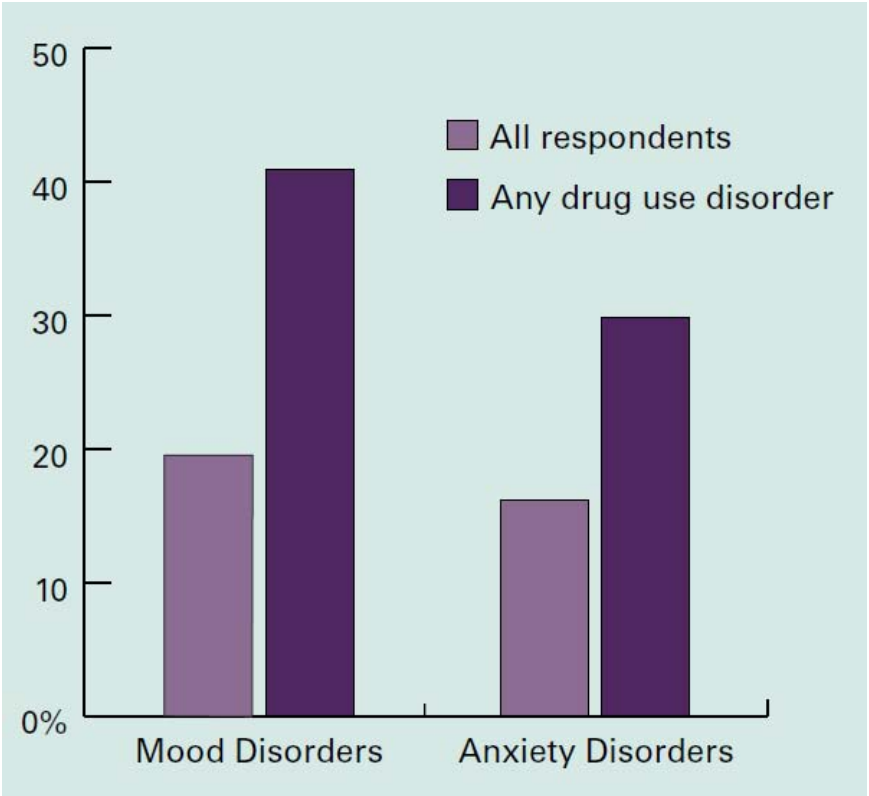
- Mental disorder $\Rightarrow$ drug use disorder
- Drug use disorder $\Rightarrow$ mental disorder
- Common etiology $\Rightarrow$ $\begin{cases} \text{mental disorder} \\ \text{drug use disorder} \end{cases}$



Source: www.nida.nih.gov

Possible Mechanisms for Comorbidity

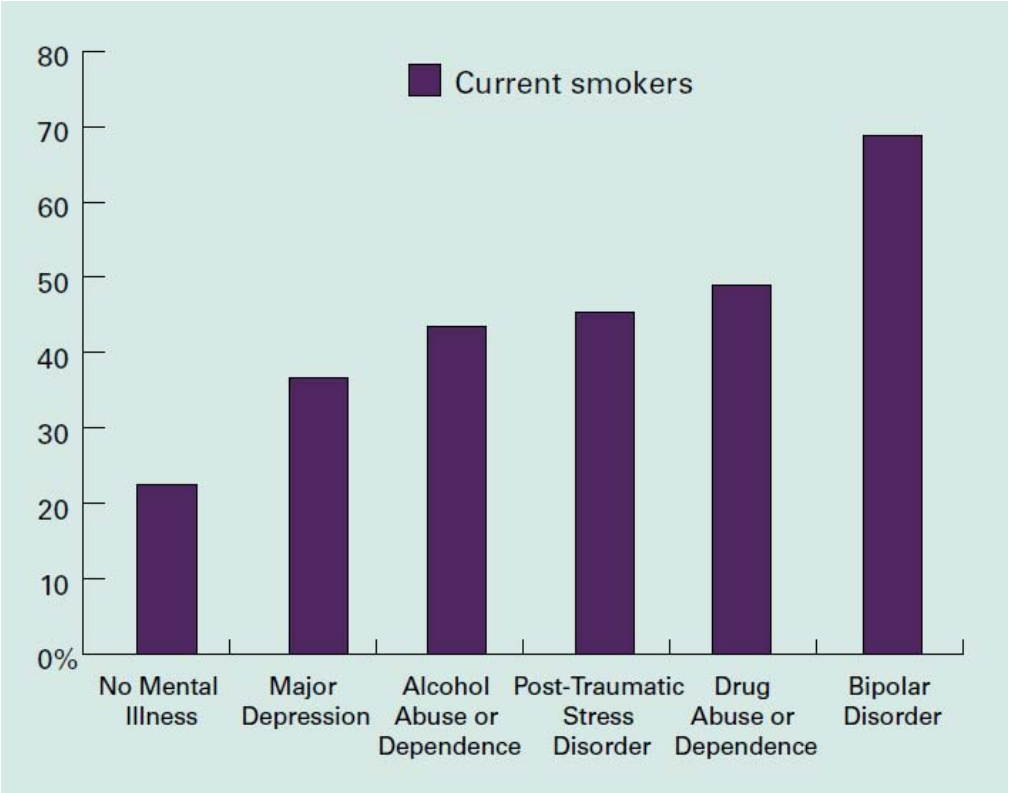
- Mental disorder $\Rightarrow$ drug use disorder
- Drug use disorder $\Rightarrow$ mental disorder
- Common etiology $\Rightarrow$ $\left\{ \begin{array}{l} \text{mental disorder} \\ \text{drug use disorder} \end{array} \right.$



Source: www.nida.nih.gov

Possible Mechanisms for Comorbidity

- Mental disorder $\Rightarrow$ drug use disorder
- Drug use disorder $\Rightarrow$ mental disorder
- Common etiology $\Rightarrow$ $\left\{ \begin{array}{l} \text{mental disorder} \\ \text{drug use disorder} \end{array} \right.$

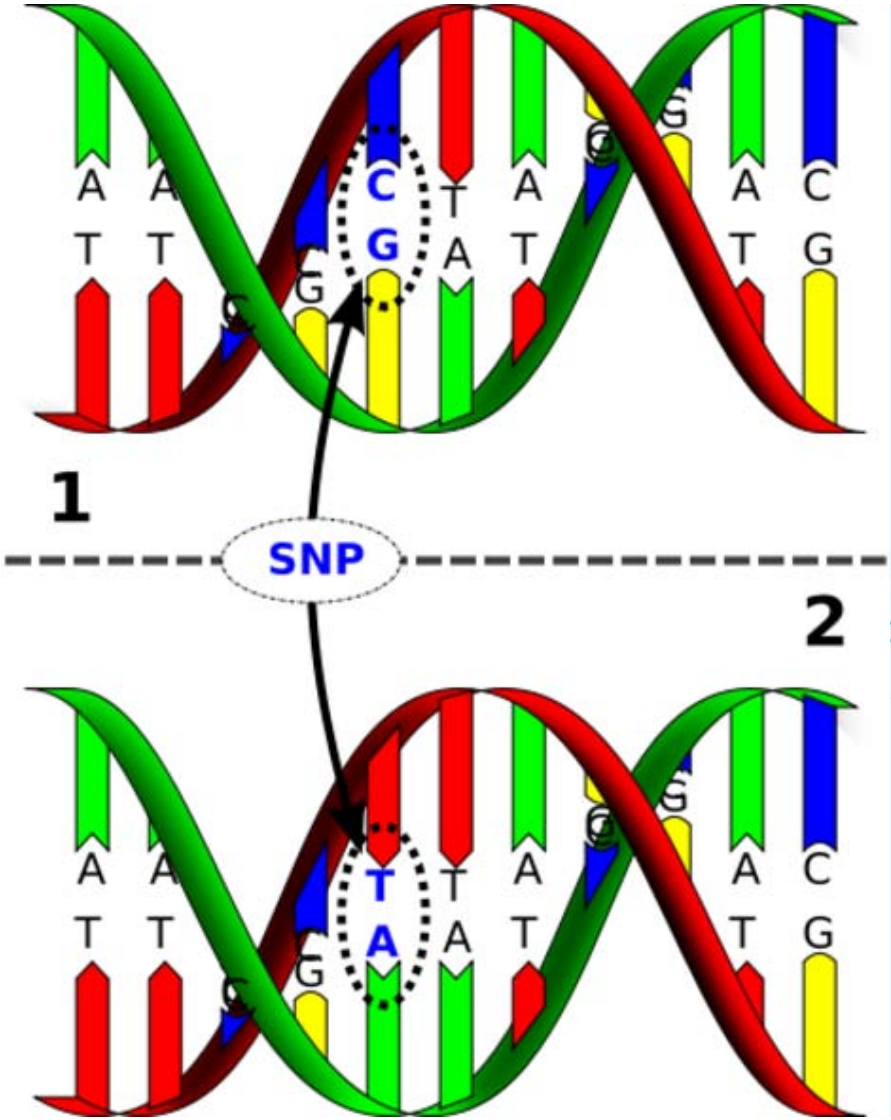


Source: www.nida.nih.gov

Outline

1. Background
   - Comorbidity
   - Disorders, Genes and Covariates
2. Weighted Association Test
   - Generalized Kendall's Tau
   - Asymptotic Distribution and Power
   - Application to WTCCC Bipolar Disorder Data
3. Maximum Weighted Association Test
   - Asymptotic Distribution
   - Simulations
   - Application to COGA Data
4. Conclusions

Genotypes and Covariates



6. **What is Person 1's sex?** *Mark ✗ ONE box.*

☐ Male ☐ Female

7. **What is Person 1's age and what is Person 1's date of birth?**

Please report babies as age 0 when the child is less than 1 year old.

Print numbers in boxes.

Age on April 1, 2010 | Month | Day | Year of birth

➜ **NOTE: Please answer BOTH Question 8 about Hispanic origin and Question 9 about race. For this census, Hispanic origins are not races.**

8. **Is Person 1 of Hispanic, Latino, or Spanish origin?**

☐ **No,** not of Hispanic, Latino, or Spanish origin
☐ Yes, Mexican, Mexican Am., Chicano
☐ Yes, Puerto Rican
☐ Yes, Cuban
☐ Yes, another Hispanic, Latino, or Spanish origin — *Print origin, for example, Argentinean, Colombian, Dominican, Nicaraguan, Salvadoran, Spaniard, and so on.*

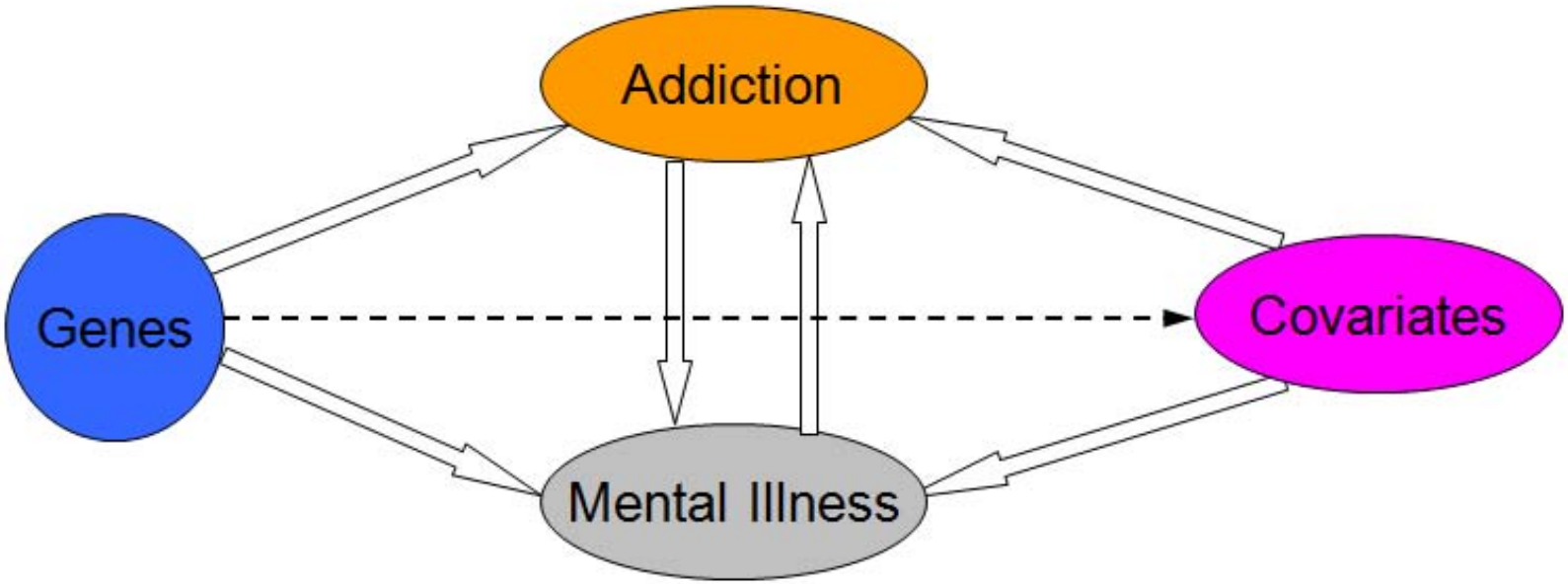
9. **What is Person 1's race?** *Mark ✗ one or more boxes.*

☐ White
☐ Black, African Am., or Negro
☐ American Indian or Alaska Native — *Print name of enrolled or principal tribe.*

☐ Asian Indian
☐ Chinese
☐ Filipino
☐ Japanese
☐ Korean
☐ Vietnamese
☐ Other Asian — *Print race, for example, Hmong, Laotian, Thai, Pakistani, Cambodian, and so on.*
☐ Native Hawaiian
☐ Guamanian or Chamorro
☐ Samoan
☐ Other Pacific Islander — *Print race, for example, Fijian, Tongan, and so on.*

Source: en.wikipedia.org; 2010.census.gov

Disorders, Genes and Covariates



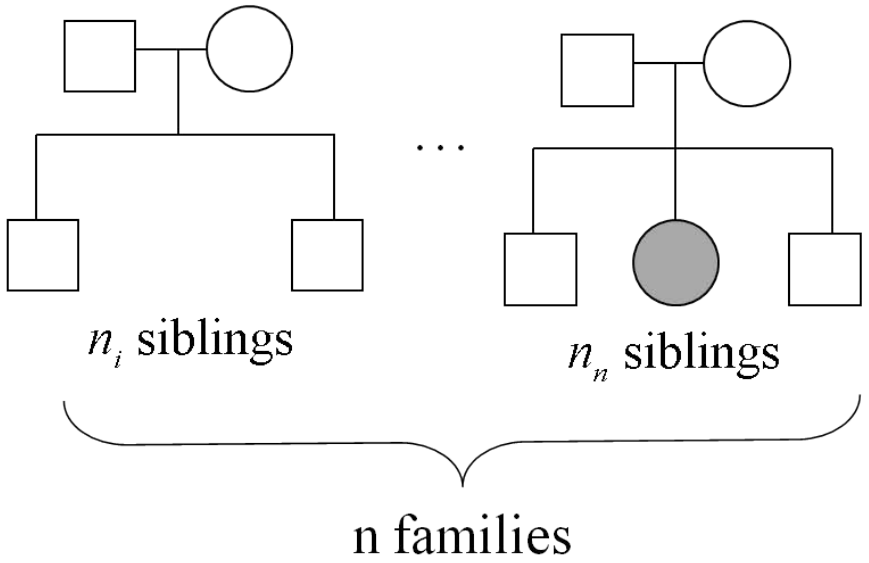
- Covariates: interact or confound genetic effects
- Failure to account for covariates: bias or reduced power

Study Design

- Population-based studies

- Family-based studies

Nuclear families



Outline

1. Background
   - Comorbidity
   - Disorders, Genes and Covariates
2. Weighted Association Test
   - Generalized Kendall's Tau
   - Asymptotic Distribution and Power
   - Application to WTCCC Bipolar Disorder Data
3. Maximum Weighted Association Test
   - Asymptotic Distribution
   - Simulations
   - Application to COGA Data
4. Conclusions

Notation and Hypothesis

- n study subjects, from a population-based study or family-based study
- For each subject:
  - A vector of traits $\mathbf{T} = (T^{(1)}, \ldots, T^{(p)})'$
  - Marker genotype M
  - Parental marker genotypes M^{pa} (only available in a family-based study)
  - A vector of covariates $\mathbf{Z} = (Z^{(1)}, \ldots, Z^{(l)})'$
- Null hypothesis: no association between marker alleles and any linked locus that influences traits $\mathbf{T}$

Typical Response

Fagerstrom Test for Nicotine Dependence

Quantitative Scale

1. How many cigarettes a day do you usually smoke?

1 to 10	0 point	21 to 30	2 points
11 to 20	1 point	30 or more	3 points

2. How soon after you wake up do you smoke your first cigarette? **Ordinal Scale**

After 60 minutes	0 point	6 - 30 minutes	2 points
31 - 60 minutes	1 point	< 5 minutes	3 points

3. Do you smoke more during the first two hours of the day than during the rest of the day?

No	0 point	Yes	1 point

4. Which cigarette would you most hate to give up?

Any other cigarette than the first one	0 point	The first cigarette in the morning	1 point

5. Do you find it difficult to refrain from smoking in places where it is forbidden, such as public buildings, on airplanes or at work?

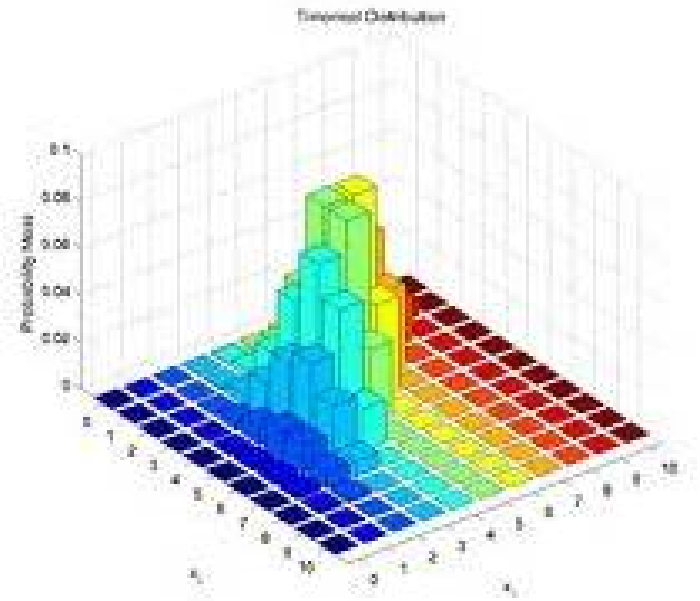
No	**Dichotomous Scale**	0 point	Yes	1 point

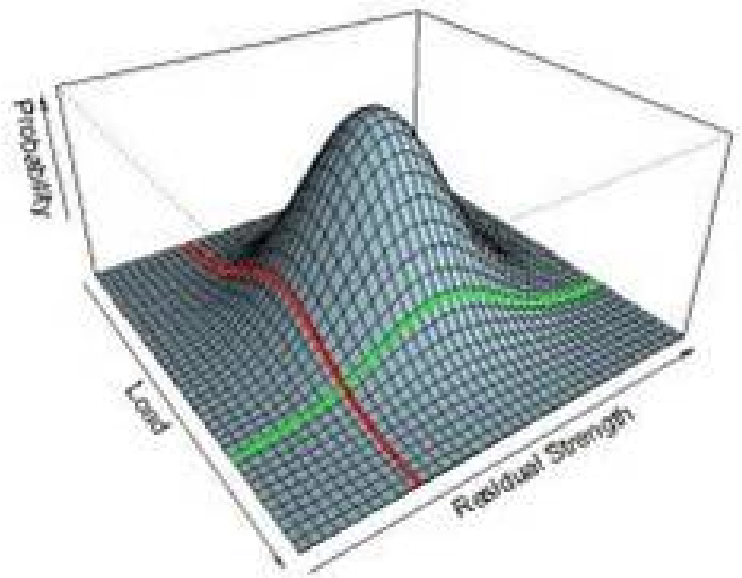
6. Do you still smoke even when you are so ill that you are in bed most of the day?

No	0 point	Yes	1 point

Total points

Multivariate Distributions





$$\prod_t \frac{n_t!}{\prod_a n_{t,a}!} \prod_a \hat{p}_{t,a}^{n_{t,a}}$$

$$\frac{\exp\left\{-\frac{1}{2}(x-\mu)'\Sigma^{-1}(x-\mu)\right\}}{\sqrt{(2\pi)^n|\Sigma|}}$$

Kendall's Tau

- A nonparametric statistic measuring the rank correlation between two variables
- Pairs of observations: $\{(X_i, Y_i) : i = 1, \ldots, n\}$
- (X_i, Y_i) and (X_j, Y_j):
  - Concordant, if $X_i - X_j$ and $Y_i - Y_j$ have the same sign
  - Disconcordant, if $X_i - X_j$ and $Y_i - Y_j$ have the different sign
- Kendall's tau:

$$\tau = 2(A - B)/\{n(n-1)\}$$

  A and B: numbers of concordant and disconcordant pairs
- Or

$$\tau = \binom{n}{2}^{-1} \sum_{i<j} \text{sign}\{(X_i - X_j)(Y_i - Y_j)\}$$

Generalized Kendall's Tau

- $\mathbf{F}_{ij} = \{f_1(T_i^{(1)} - T_j^{(1)}), \ldots, f_p(T_i^{(p)} - T_j^{(p)})\}'$
  - $f_k(\cdot)$: identity function for a quantitative or binary trait
  - $f_k(\cdot)$: sign function for an ordinal trait
- $D_{ij} = C_i - C_j$. C: number of any chosen allele in marker genotype M
- Genaralized Kendall's tau (Zhang, Liu and Wang, 2010):

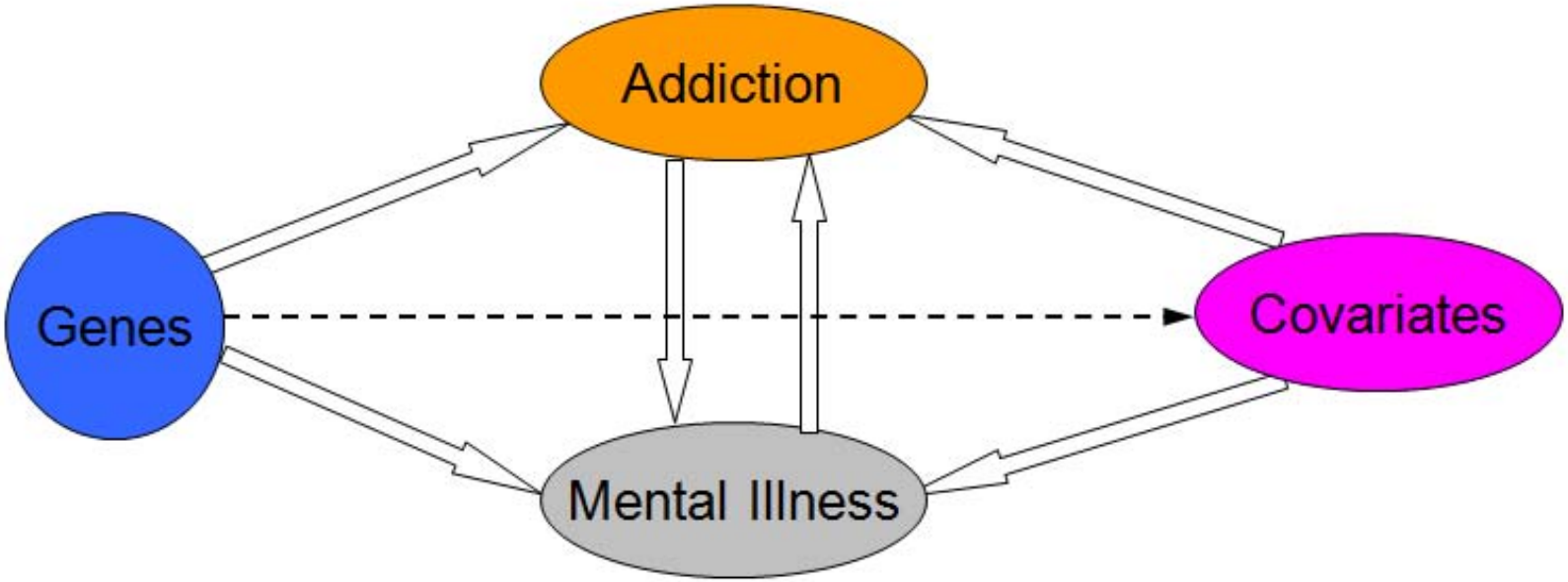
$$\mathbf{U} = \binom{n}{2}^{-1} \sum_{i<j} D_{ij}\mathbf{F}_{ij}$$

- Special cases:
  - FBAT-GEE (Lange et al. 2003)
  - Test for a single ordinal trait (Wang, Ye and Zhang, 2006)

Outline

1. Background
   - Comorbidity
   - Disorders, Genes and Covariates
2. Weighted Association Test
   - Generalized Kendall's Tau
   - Asymptotic Distribution and Power
   - Application to WTCCC Bipolar Disorder Data
3. Maximum Weighted Association Test
   - Asymptotic Distribution
   - Simulations
   - Application to COGA Data
4. Conclusions

Hypothesis with Covariates



- New null hypothesis: no association between marker alleles and any linked locus that influences traits $\mathbf{T}$ conditional on covariates $\mathbf{Z}$

Weighted Test

- A weight function $w(\mathbf{Z}_i, \mathbf{Z}_j)$ imposes a relatively large weight when $\mathbf{Z}_i$ is close to $\mathbf{Z}_j$, and a relatively small weight when $\mathbf{Z}_i$ and $\mathbf{Z}_j$ are far away
- Weighted U-statistic:

$$\mathbf{S} = \binom{n}{2}^{-1} \sum_{i<j} D_{ij}\mathbf{F}_{ij} w(\mathbf{Z}_i, \mathbf{Z}_j)$$

- Weighted test statistic:

$$\chi^2_\tau = \{\mathbf{S} - E_0(\mathbf{S})\}' \mathrm{Var}_0^{-1}(\mathbf{S}) \{\mathbf{S} - E_0(\mathbf{S})\}$$

Weight Function–I: Distance

- Write $\mathbf{Z} = (\mathbf{Z}^{co\prime}, \mathbf{Z}^{ca\prime})'$, with $\mathbf{Z}^{co}$ for the continuous covariates and $\mathbf{Z}^{ca}$ for the categorical covariates

$$w(\mathbf{Z}_i, \mathbf{Z}_j; h, q) = W_h(\|\mathbf{Z}_i^{co} - \mathbf{Z}_j^{co}\|) W_q\{I(\mathbf{Z}_i^{ca} \neq \mathbf{Z}_j^{ca})\}$$

- For example,

$$W_h(u) = \exp(-u^2/2h^2), \ h > 0,$$

$$W_q(v) = (1-q)I(v=0) + qI(v=1), \ 0 \leq q \leq 0.5$$

- Weighted U-statistic (called fixed-(h, q) U-statistic):

$$\mathbf{S}(h, q) = \binom{n}{2}^{-1} \sum_{i<j} D_{ij} \mathbf{F}_{ij} w(\mathbf{Z}_i, \mathbf{Z}_j; h, q)$$

Weight Function–II: Propensity Score

- Propensity score: probability of a unit being assigned to a particular treatment given a set of covariates
- Causal effect analysis: match subjects according to their propensity scores (Rosenbaum and Rubin, 1984)
- Genomic propensity score: $p(\mathbf{z}) = \{p_1(\mathbf{z}), p_2(\mathbf{z})\}'$, $p_c(\mathbf{z}) = P(C = c|\mathbf{Z} = \mathbf{z})$
- Genetic association analysis: match subjects according to their genomic propensity scores
- Weight function:

$$w(\mathbf{Z}_i, \mathbf{Z}_j) = W_h\{\|p(\mathbf{Z}_i) - p(\mathbf{Z}_j)\|\},$$

with $W_h(u) = \exp(-u^2/2h^2),\ h > 0$

Asymptotic Distribution: Setting

- Treating the offspring genotype as random
- Conditioning on all phenotypes and parental genotypes (if available)
- Eliminates the assumptions about phenotype distribution, genetic model and parental genotype distribution
- Robust and less prone to population stratification
- In addition, conditioning on covariates

Asymptotic Distribution: Null Hypothesis

- When $n \to \infty$,

$$\mathrm{Var}_0^{-1/2}\{\mathbf{S}(h,q)\}[\mathbf{S}(h,q) - E_0\{\mathbf{S}(h,q)\}] \xrightarrow{\mathcal{D}} N(\mathbf{0}, \mathbf{I}_p)$$

- Fixed-(h,q) test statistic:

$$\chi^2_\tau(h,q) \xrightarrow{\mathcal{D}} \chi^2_p$$

- Mean and variance:

$$E_0\{\mathbf{S}(h,q)\} = \frac{2}{n-1}\sum_{i=1}^{n} \bar{\mathbf{u}}_i E_0(C_i | M_i^{pa}, \mathbf{Z}_i),$$

$$\mathrm{Var}_0\{\mathbf{S}(h,q)\} = \frac{4}{(n-1)^2}\sum_{i=1}^{n}\sum_{i=1}^{n} \bar{\mathbf{u}}_i \bar{\mathbf{u}}_j' \mathrm{Cov}_0(C_i, C_j | M_i^{pa}, M_j^{pa}, \mathbf{Z}_i, \mathbf{Z}_j),$$

with $\bar{\mathbf{u}}_i = n^{-1}\sum_{j=1}^{n} \mathbf{F}_{ij} w(\mathbf{Z}_i, \mathbf{Z}_j; h, q)$

Asymptotic Distribution: Power

- Under the alternative hypothesis,

$$\chi^2_\tau(h,q) \sim \sum_{i=1}^{p} e_i \chi^2_1(\phi_i)$$

  - $\Delta\boldsymbol{\mu} = \boldsymbol{\mu}_1 - \boldsymbol{\mu}_0 \equiv E_1\{\mathbf{S}(h,q)\} - E_0\{\mathbf{S}(h,q)\}$
  - $\boldsymbol{\Sigma}_0 = \mathrm{Var}_0\{\mathbf{S}(h,q)\}$
  - $\boldsymbol{\Sigma}_1 = \mathrm{Var}_1\{\mathbf{S}(h,q)\}$
  - $e_1 \geq \cdots \geq e_p \geq 0$: eigenvalues of $\boldsymbol{\Sigma}_1^{1/2}\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\Sigma}_1^{1/2}$
  - $\phi_i = \Delta\tilde{\mu}_i^2$
  - $\Delta\tilde{\mu}_i$: ith component of $\Delta\tilde{\boldsymbol{\mu}} = \mathbf{Q}\boldsymbol{\Sigma}_1^{-1/2}\Delta\boldsymbol{\mu}$
  - $\mathbf{Q}$: an orthonormal matrix, $\mathbf{Q}\boldsymbol{\Sigma}_1^{1/2}\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\Sigma}_1^{1/2}\mathbf{Q}' = \mathrm{diag}(e_1,\ldots,e_p)$

Factors Determining the Power

- The conditional power $\mathcal{P}$: $\mathcal{P} = P\left\{\sum_{i=1}^{p} e_i \chi_1^2(\phi_i) \geq q_{\chi_p^2}(1-\alpha)\right\}$
- Taking a family-based study as an example,

$$\boldsymbol{\mu}_1 = \frac{2}{n-1}\sum_{i=1}^{n} \bar{\mathbf{u}}_i E(C_i|\mathbf{T}_i, \mathbf{Z}_i, M_i^{pa})$$

$$\boldsymbol{\Sigma}_1 = \frac{4}{(n-1)^2}\sum_{i=1}^{n}\sum_{j=1}^{n} \bar{\mathbf{u}}_i \bar{\mathbf{u}}_j' \mathrm{Cov}(C_i, C_j|\mathbf{T}_i, \mathbf{T}_j, \mathbf{Z}_i, \mathbf{Z}_j, M_i^{pa}, M_j^{pa})$$

- By Bayes' theorem, $P(C=c|\mathbf{T}, \mathbf{Z}, M^{pa}) = \frac{P(\mathbf{T}|C=c,\mathbf{Z})P(C=c|M^{pa})}{\sum_{c'} P(\mathbf{T}|C=c',\mathbf{Z})P(C=c'|M^{pa})}$
  - Penetrance: $P(\mathbf{T}|C=c, \mathbf{Z})$
  - Allele frequency: $P(C=c|M^{pa})$

Power Approximation

Using the result from Liu et al. (2009), we have

Theorem

$$\mathcal{P} \approx P\{\chi^2_l(\nu) \geq q^*\},$$

where l, ν, *and* q^* *depend on* $\boldsymbol{\mu}_1$ *and* $\boldsymbol{\Sigma}_1$.

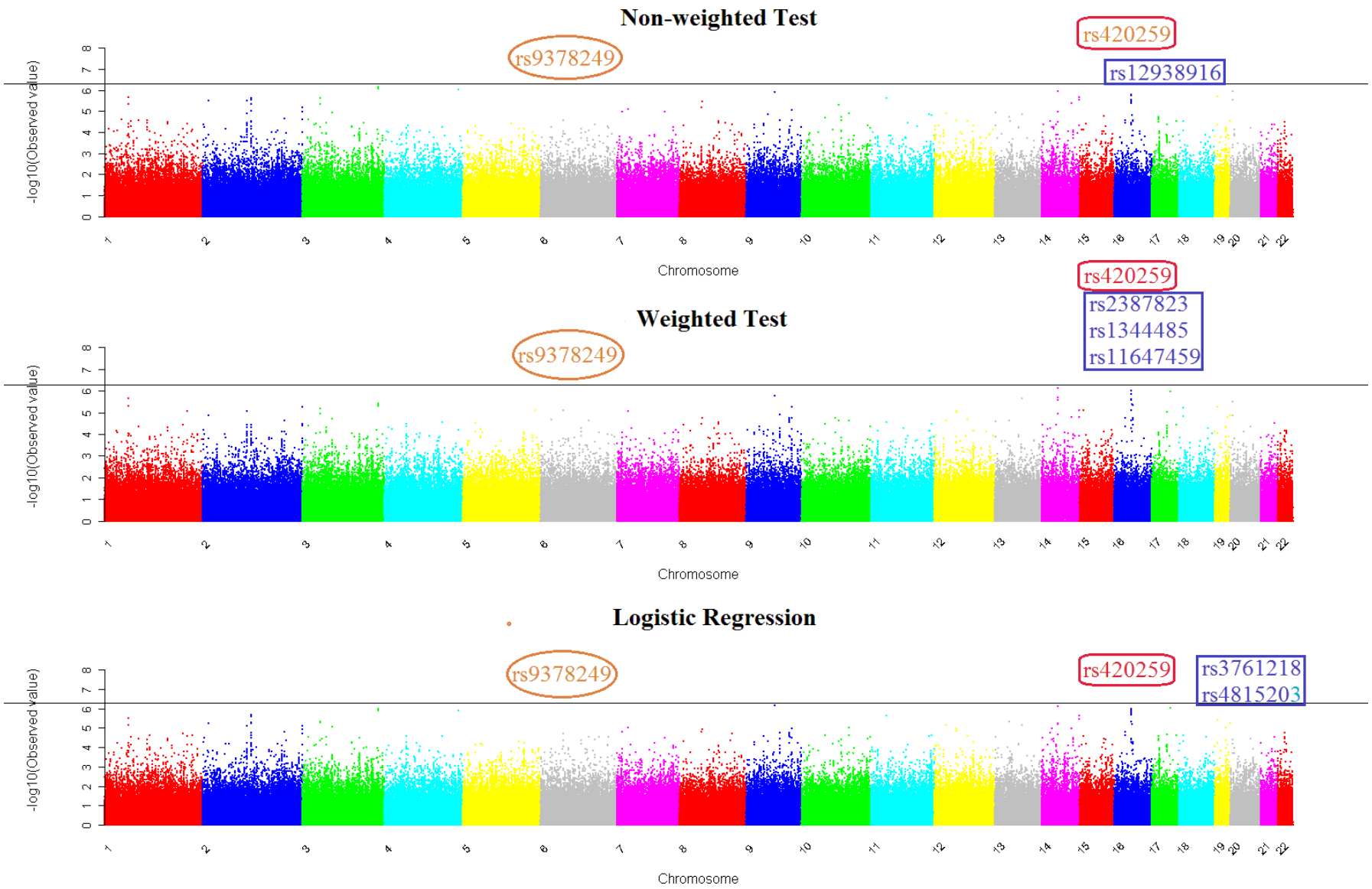
Outline

1. Background
   - Comorbidity
   - Disorders, Genes and Covariates
2. Weighted Association Test
   - Generalized Kendall's Tau
   - Asymptotic Distribution and Power
   - Application to WTCCC Bipolar Disorder Data
3. Maximum Weighted Association Test
   - Asymptotic Distribution
   - Simulations
   - Application to COGA Data
4. Conclusions

WTCCC Bipolar Disorder Data

- Collected by Wellcome Trust Case-Control Consortium (WTCCC, 2007, Nature)
  - Phenotype: 1998 cases/3004 controls of bipolar disorder
  - Genotype: genotyped by Affymetrix GeneChip 500K arrays
  - Covariates: gender, age at recruitment
- Our method: weighted test using propensity score approach ($h = 1$)
- Methods for comparison: non-weighted test and logistic regression
- Strong association: p-value $< 5 \times 10^{-7}$; moderate association: $5 \times 10^{-7} <$ p-value $< 10^{-5}$

Manhattan Plot: Comparison of Three Methods



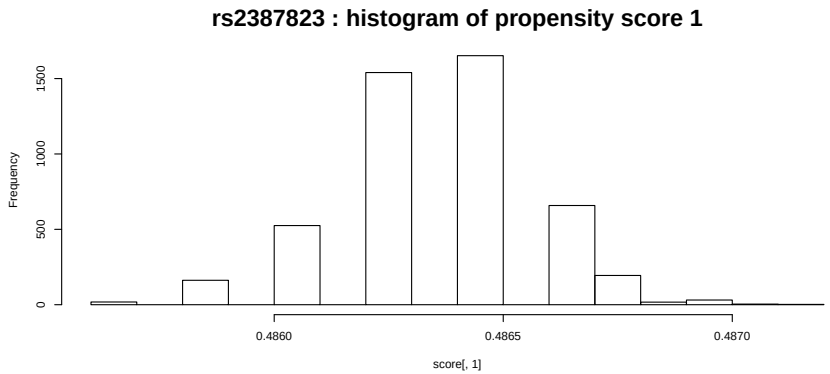
GWAS Results

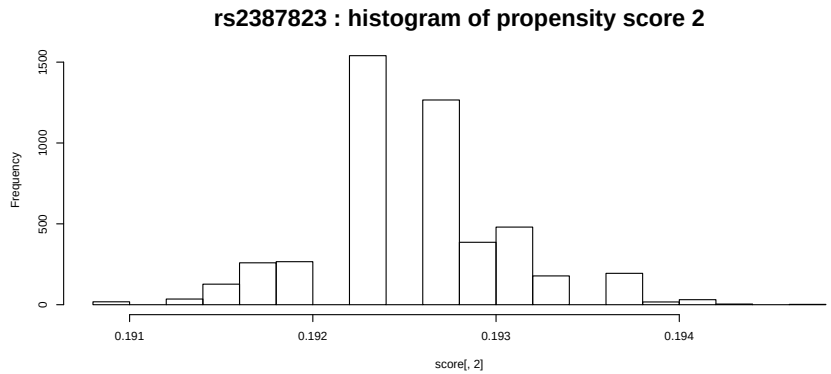
Chr.	SNP	Position	Non-weighted	Weighted	Logistic Regression
6	rs9378249	31435680	1.21e-8	1.39e-8	1.71e-9
16	rs420259	23541527	8.51e-9	6.59e-8	3.33e-9
16	rs2387823	51445620	2.90e-6	1.30e-7	1.77e-6
16	rs1344485	51469833	1.78e-6	1.79e-7	1.41e-6
16	rs11647459	51473252	2.93e-6	2.76e-7	1.89e-6
17	rs12938916	53221286	4.80e-7	1.11e-6	8.89e-7
20	rs4815603	3720527	3.00e-6	1.42e-5	4.80e-7
20	rs37612181	3724175	1.13e-6	3.27e-6	2.16e-7

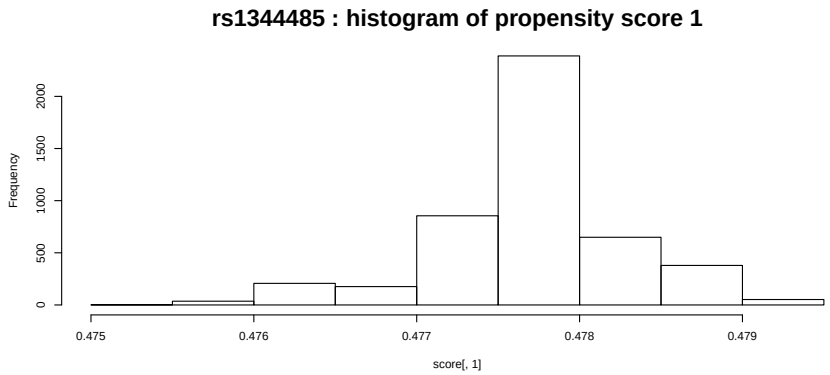
Propensity Scores: Estimation

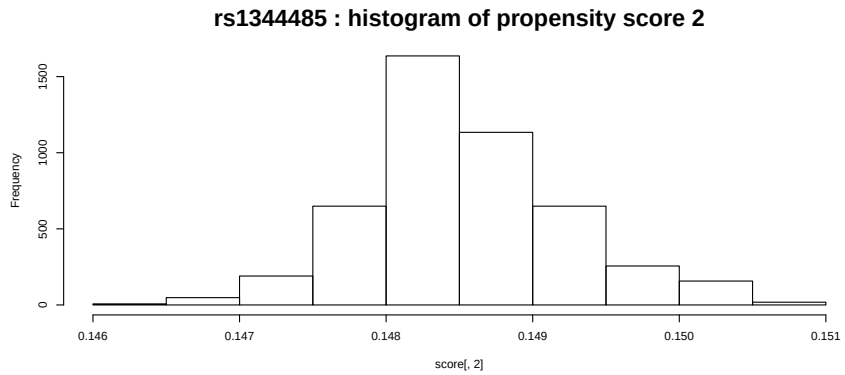
Chr.	SNP	Gender		Age	
		Coefficient	p-value	Coefficient	p-value
16	rs2387823	0.0021	0.970	0.0031	0.902
16	rs1344485	-0.0028	0.959	-0.0049	0.850
16	rs11647459	0.0004	0.994	0.0038	0.882

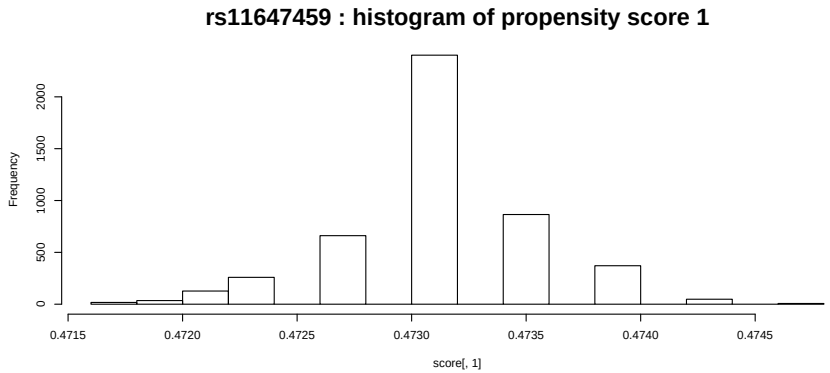
Propensity Scores: Histograms

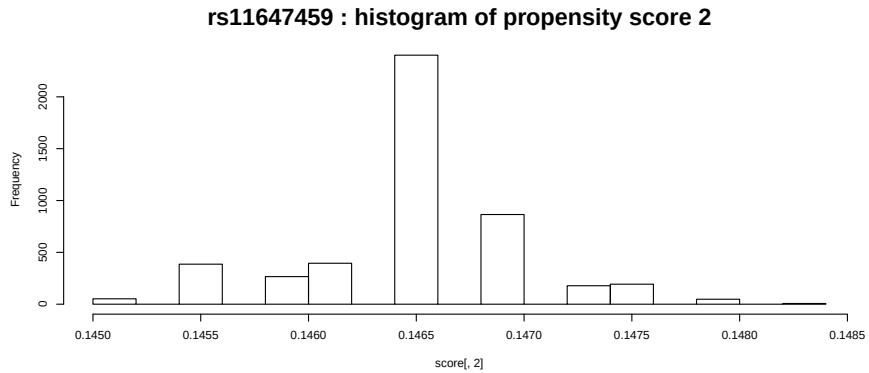




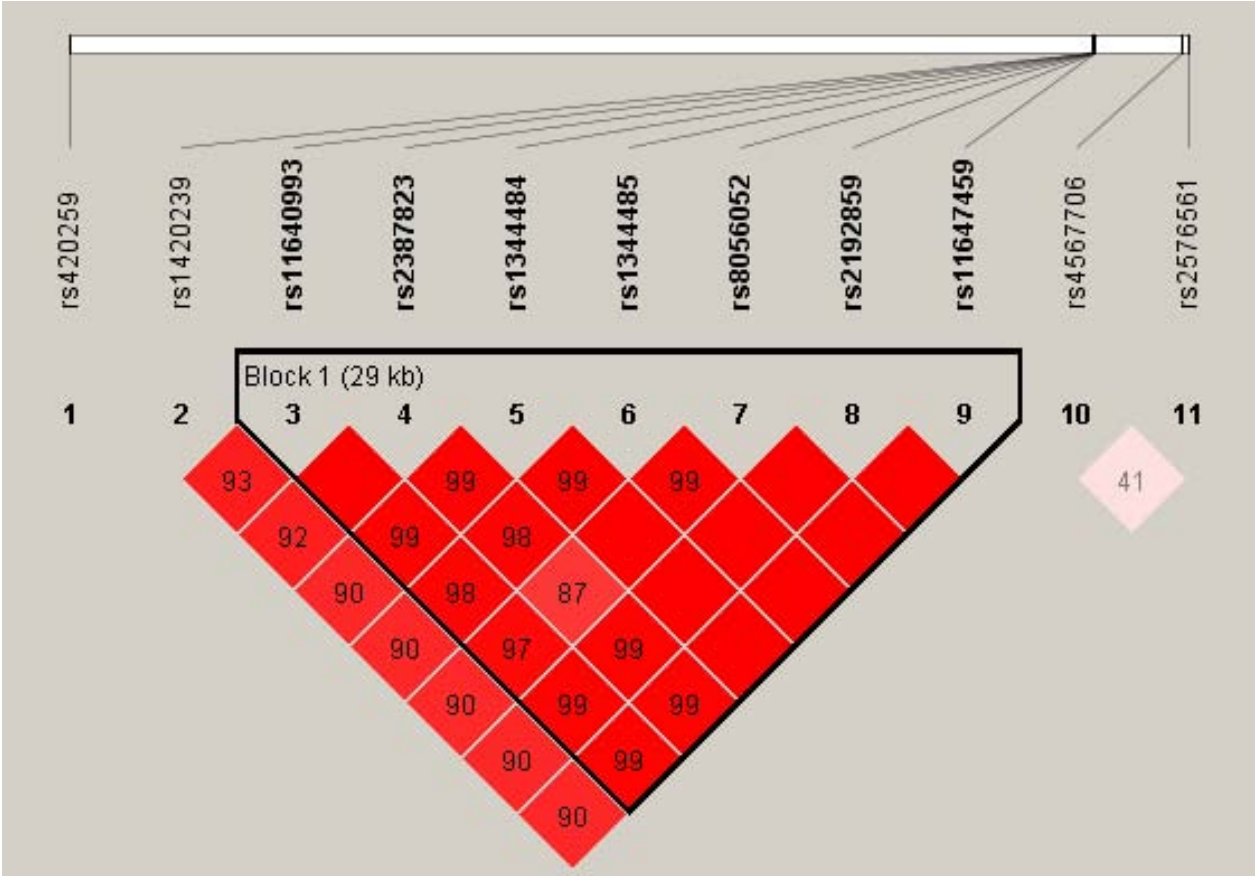








Significant Region



- 29kb region: 7 strongly linked SNPs
- Haplotype association p-value: 2.64×10^{-7} by weighted test

Interactions Between rs420259 and New SNPs

Model	Variable	Coefficient	p-value
(1)	rs420259	-0.77415	6.43e-8
	rs9378249	-0.60212	3.88e-8
	rs420259 $\times$ rs9378249	0.37602	0.332
(2)	rs420259	-0.63671	1.51e-3
	rs2387823	-0.18834	1.73e-5
	rs420259 $\times$ rs2387823	-0.10661	0.561
(3)	rs420259	-0.62866	1.10e-3
	rs1344485	-0.19898	1.14e-5
	rs420259 $\times$ rs1344485	-0.10653	0.575
(4)	rs420259	-0.67070	4.43e-4
	rs11647459	-0.19508	1.53e-5
	rs420259 $\times$ rs11647459	-0.07443	0.693

Interactions Among New SNPs

Model	Variable	Coefficient	p-value
(1)	rs9378249	-0.49915	1.77e-3
	rs2387823	-0.18455	4.62e-5
	rs9378249 $\times$ 2387823	-0.10876	0.461
(2)	rs9378249	-0.49551	1.07e-3
	rs1344485	-0.20266	1.58e-5
	rs9378249 $\times$ 1344485	-0.14683	0.344
(3)	rs9378249	-0.48910	1.13e-3
	rs11647459	-0.19580	2.74e-5
	rs9378249 $\times$ rs11647459	-0.14960	0.333

Outline

1. Background
   - Comorbidity
   - Disorders, Genes and Covariates
2. Weighted Association Test
   - Generalized Kendall's Tau
   - Asymptotic Distribution and Power
   - Application to WTCCC Bipolar Disorder Data
3. Maximum Weighted Association Test
   - Asymptotic Distribution
   - Simulations
   - Application to COGA Data
4. Conclusions

Maximum Weighted Test

- Fixed-(h, q) test: how to choose optimal parameters h and q?
- Choose a grid of h and q values and maximize the weighted test statistic over those choices
- $\{h_1, \ldots, h_{L_1}\}$: pre-specified grid points of h
- $\{q_1, \ldots, q_{L_2}\}$: pre-specified grid points of q

$$\chi^2_{\tau,\max} = \max_{1 \le l_1 \le L_1, 1 \le l_2 \le L_2} \chi^2_\tau(h_{l_1}, q_{l_2})$$

- Approximate the optimal weighting scheme, yielding the strongest association measure

Resampling Approach

- Population-based studies: restricted permutation in Yu et al. (2010)
- Family-based studies: children's genotypes solely determined by their parents' marker alleles, resample the children's genotype by Mendelian laws
- Calculate M resampling test statistics $\tilde{\chi}^2_{\tau,\max,1}, \ldots, \tilde{\chi}^2_{\tau,\max,M}$ using M resampled data
- Resampling p-value: the proportion of the resampling test statistics that exceed our observed test statistic, i.e., $M^{-1} \sum_{m=1}^{M} I(\tilde{\chi}^2_{\tau,\max,m} \geq \chi^2_{\tau,\max})$

Asymptotic Distribution: Joint Distribution

- Equivalently,

$$\chi^2_{\tau,\max} = \max_{1\le l_1\le L_1, 1\le l_2\le L_2} \|\mathbf{R}_{l_1,l_2}\|^2$$

  - $\mathbf{R} = \mathrm{Var}_{0D}^{-1/2}(\mathbf{S})\{\mathbf{S} - E_0(\mathbf{S})\}$
  - $\mathbf{S} = \{\mathbf{S}'(h_1, q_1), \ldots, \mathbf{S}'(h_{L_1}, q_{L_2})\}'$
  - $\mathrm{Var}_{0D}(\mathbf{S}) = \mathrm{diag}[\mathrm{Var}_0\{\mathbf{S}(h_1, q_1)\}, \ldots, \mathrm{Var}_0\{\mathbf{S}(h_{L_1}, q_{L_2})\}]$: the diagonal blocks of $\mathrm{Var}_0(\mathbf{S})$

- $$\mathrm{Var}_0^{-1/2}(\mathbf{S})\{\mathbf{S} - E_0(\mathbf{S})\} \xrightarrow{\mathcal{D}} N(\mathbf{0}, \mathbf{I}_{pL_1L_2})$$

- $\tilde{\mathbf{R}} = \mathrm{Var}_{0D}^{-1/2}(\mathbf{S})\mathrm{Var}_0^{1/2}(\mathbf{S})\mathbf{G}$, $\mathbf{G} \sim N(\mathbf{0}, \mathbf{I}_{pL_1L_2})$

Asymptotic Distribution: Uniform Approximation

Theorem

Assume that the eigenvalues of $\mathrm{Var}_{0D}(\mathbf{S})$ *and* $\mathrm{Var}_0(\mathbf{S})$ *are uniformly bounded from both above and below, i.e., there exist two positive numbers* c *and* C *such that* $c \leq \lambda_{\min}\{\mathrm{Var}_{0D}(\mathbf{S})\} \leq \lambda_{\max}\{\mathrm{Var}_{0D}(\mathbf{S})\} \leq C$ *and* $c \leq \lambda_{\min}\{\mathrm{Var}_0(\mathbf{S})\} \leq \lambda_{\max}\{\mathrm{Var}_0(\mathbf{S})\} \leq C$ *uniformly for all* n*, where* $\lambda_{\min}$ *and* $\lambda_{\max}$ *denote the smallest and largest eigenvalues respectively. Then for any* $x \in \mathbb{R}$*, as* $n \to \infty$*,*

$$\sup_{x\in\mathbb{R}} \left| P\left(\chi^2_{\tau,\max} \leq x\right) - P\left(\max_{1\leq l_1\leq L_1, 1\leq l_2\leq L_2} \|\tilde{\mathbf{R}}_{l_1,l_2}\|^2 \leq x\right)\right| \to 0.$$

Outline

1. Background
   - Comorbidity
   - Disorders, Genes and Covariates
2. Weighted Association Test
   - Generalized Kendall's Tau
   - Asymptotic Distribution and Power
   - Application to WTCCC Bipolar Disorder Data
3. Maximum Weighted Association Test
   - Asymptotic Distribution
   - Simulations
   - Application to COGA Data
4. Conclusions

Aims

- Compare the performance of:
  - Maximum weighted test
  - Non-weighted test
- Compare the performance of:
  - Maximum weighted test
  - Other covariate-adjusted tests

Simulation I: Data Generation

- Generate the parents' disease and marker genotypes via the haplotype frequencies
- Given the parental genotypes, generate the offspring genotype using 1cM between the two loci
- Two covariates are generated: $Z^{co} \sim N(1, 2)$
  - Without confounder: $P(Z^{ca} = 1) = 1 - P(Z^{ca} = 0) = 0.7$
  - With confounder: $\text{logit}\{P(Z^{ca} = 1)\} = 0.5M^{fa} + 0.5M^{mo}$
- Bivariate ordinal traits are generated according to random effects proportional odds model:

$$\text{logit}\{P(T^{(j)} \leq k)\} = \alpha_{j,k} + \beta_g G + \beta_{co} Z^{co} + \beta_{ca} Z^{ca} + U_j,$$

with $k = 1, \ldots, K_j$, $j = 1, 2$, and $(U_1, U_2) \sim N(0, \Sigma)$

Simulation I: Parameters

- Number of categories: $K_1 = 3$ and $K_2 = 4$
- $(\alpha_{1,1}, \alpha_{1,2}) = (-0.5, -0.3)$, $(\alpha_{2,1}, \alpha_{2,2}, \alpha_{2,3}) = (-0.5, -0.3, -0.1)$
- $\beta_g = 2.0$
- $\beta_{co} = \beta_{ca} = 0, 0.5, 1.0, 1.5, 2.0$
- $\mathbf{\Sigma} = \begin{pmatrix} 1 & 0.25 \\ 0.25 & 1 \end{pmatrix}$
- The grid of h is $\{C_1(C_2/C_1)^{(l_1/(L_1-1))} : \ l_1 = 0, \ldots, L_1 - 1\}$, with $C_1 = 0.05$, $C_2 = 10$, $L_1 = 8$
- The grid of q is $\{0.5 l_2/(L_2 - 1) : \ l_2 = 0, \ldots, L_2 - 1\}$, with $L_2 = 5$

Type I Error of Maximum Weighted Test

Confounder	n	Significance Level		
		$\alpha = 0.05$	$\alpha = 0.01$	$\alpha = 0.001$
No	200	0.0466	0.0090	0.0006
	400	0.0512	0.0097	0.0010
Yes	200	0.0469	0.0084	0.0007
	400	0.0462	0.0090	0.0011

Power Comparison: Without Confounder

n	α	Method	Covariate effect 0.0	0.5	1.0	1.5	2.0
200	0.05	Weighted	0.681	0.521	0.372	0.275	0.222
		Non-weighted	0.726	0.522	0.306	0.189	0.135
	0.01	Weighted	0.432	0.281	0.161	0.099	0.071
		Non-weighted	0.491	0.283	0.128	0.064	0.041
	0.001	Weighted	0.160	0.082	0.036	0.017	0.011
		Non-weighted	0.223	0.097	0.028	0.011	0.006
400	0.05	Weighted	0.948	0.848	0.685	0.551	0.448
		Non-weighted	0.960	0.838	0.565	0.348	0.233
	0.01	Weighted	0.846	0.658	0.441	0.297	0.213
		Non-weighted	0.877	0.643	0.321	0.154	0.084
	0.001	Weighted	0.563	0.337	0.164	0.091	0.054
		Non-weighted	0.671	0.361	0.115	0.040	0.018

Power Comparison: With Confounder

n	α	Method	Covariate effect 0.0	0.5	1.0	1.5	2.0
200	0.05	Weighted	0.695	0.557	0.405	0.310	0.250
		Non-weighted	0.728	0.528	0.308	0.194	0.139
	0.01	Weighted	0.452	0.305	0.181	0.120	0.087
		Non-weighted	0.495	0.288	0.129	0.066	0.041
	0.001	Weighted	0.165	0.090	0.046	0.024	0.015
		Non-weighted	0.224	0.095	0.032	0.011	0.006
400	0.05	Weighted	0.951	0.867	0.718	0.593	0.493
		Non-weighted	0.961	0.834	0.573	0.363	0.251
	0.01	Weighted	0.854	0.682	0.483	0.345	0.250
		Non-weighted	0.875	0.645	0.332	0.170	0.094
	0.001	Weighted	0.572	0.355	0.196	0.111	0.069
		Non-weighted	0.665	0.364	0.129	0.044	0.020

Simulation II: Other Covariate-Adjusted Methods

- FBAT-GEE (Lange et al. 2003) adjusting for covariates:
  - Fit the regression model $g(E[T^{(j)}]) = \alpha_j + \lambda_j' \mathbf{Z}$, with $g(\cdot)$ an appropriate link function
  - Replace the original traits $T^{(j)}$ with the residuals $T^{(j)} - g^{-1}(\alpha_j + \lambda_j' \mathbf{Z})$ in the FBAT-GEE test statistic
- Ordinal trait test (Wang, Ye and Zhang, 2006):
  - Deal with a single ordinal trait at a time
  - Apply the Bonferroni correction for multiple trait testing

Simulation II: Data Generation

- Continuous covariate: $Z^{co} \sim N(1, 2)$
- Bivariate quantitative traits:

$$Y^{(j)} = \mu + \beta_g G + \beta_{co} Z^{co} + \epsilon_j,\ j = 1, 2,$$

  with $(\epsilon_1, \epsilon_2)' \sim 2\phi_2(\mathbf{x}; \mathbf{\Sigma})\Phi(\boldsymbol{\alpha}'\mathbf{x})$, $\mathbf{x} \in \mathbb{R}^2$ (bivariate skew normal distribution)
- Bivariate ordinal traits: discretizing quantitative traits by (50%, 67%) and (33%, 54%, 75%) percentiles
- $\alpha = (5, 5)$ and $\mathbf{\Sigma} = \begin{pmatrix} 1 & 0.25 \\ 0.25 & 1 \end{pmatrix}$
- $\mu = 0, \beta_g = \beta_{co} = 0.8$

Power Comparison

n	Method	Significance Level		
		$\alpha = 0.05$	$\alpha = 0.01$	$\alpha = 0.001$
200	$\chi^2_{\tau,\max}$	0.640	0.381	0.134
	FBAT-GEE	0.608	0.355	0.137
	Wang et al.'s Test	0.448	0.236	0.081
400	$\chi^2_{\tau,\max}$	0.930	0.815	0.518
	FBAT-GEE	0.902	0.758	0.499
	Wang et al.'s Test	0.775	0.590	0.320
600	$\chi^2_{\tau,\max}$	0.991	0.961	0.817
	FBAT-GEE	0.982	0.938	0.787
	Wang et al.'s Test	0.925	0.807	0.585

Outline

1. Background
   - Comorbidity
   - Disorders, Genes and Covariates
2. Weighted Association Test
   - Generalized Kendall's Tau
   - Asymptotic Distribution and Power
   - Application to WTCCC Bipolar Disorder Data
3. Maximum Weighted Association Test
   - Asymptotic Distribution
   - Simulations
   - Application to COGA Data
4. Conclusions

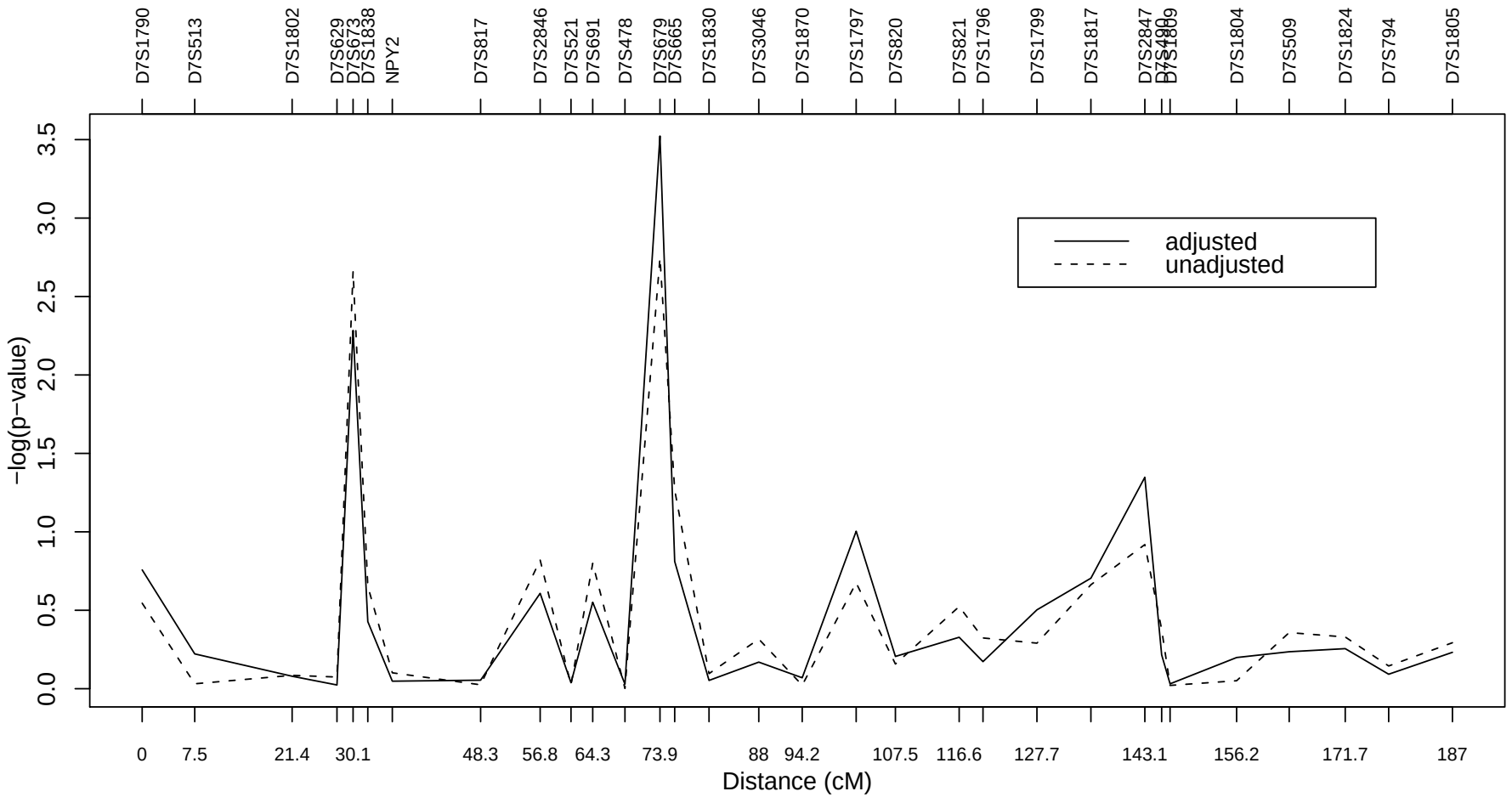
Collaborative Studies on Genetics of Alcoholism

A large scale study to map alcohol dependence susceptible genes

COGA Data

- The data include 143 families with a total of 1,614 individuals
- Multiple Traits:
  - ALDX1 (the severity of the alcohol dependence): pure unaffected, never drunk, unaffected with some symptoms, and affected
  - MaxDrink (maximum number of drinks in a 24 hour period): 0-9, 10-19, 20-29, and more than 30 drinks
  - TimeDrink (spent so much time drinking, had little time for anything else): “no”, “yes and lasted less than a month”, and “yes and lasted for one month or longer”
- Genotypes: markers on chromosome 7
- Covariates: age at interview and gender

Results


D7S1790
D7S513
D7S1802
D7S629
D7S673
D7S1838
NPY2
D7S817
D7S2846
D7S521
D7S691
D7S478
D7S679
D7S665
D7S1830
D7S3046
D7S1870
D7S1797
D7S820
D7S821
D7S1796
D7S1799
D7S1817
D7S2847
D7S1804
D7S509
D7S1824
D7S794
D7S1805
adjusted
unadjusted
−log(p−value)
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
0
7.5
21.4
30.1
48.3
56.8
64.3
73.9
88
94.2
107.5
116.6
127.7
143.1
156.2
171.7
187
Distance (cM)

Conclusions: Method

- Developed a nonparametric weighted test to adjust for covariates that accommodates multiple traits
- Provided its asymptotic distribution and analytical power calculation
- Refined the weighted test by proposing the idea of maximum weighting over the grid points of parameters
- Proposed an asymptotic approach to assess its significance

Conclusions: Application

- WTCCC bipolar disorder data: not only confirmed the results reported by the WTCCC (2007), but also identified another region at the genome-wide significance level
- The identified haplotype block is near the RPGRIP1L gene that was reported to be associated with bipolar disorder (O'Donovan et al., 2008; Riley et al., 2009)
- COGA data: confirmed and strengthened the top signal; provided evidences for the advantage of maximum weighted test over non-weighted test

Other Ongoing/Future Work

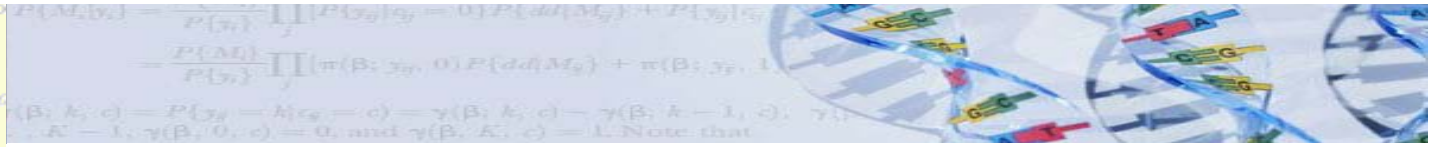
- Incorporating genetic prior information into a current study
- Genetic association analysis for rare variants
- Nonparametric test for gene-environment interactions
- Genetic test for multiple trait covariance structure

Acknowledgment

- Dr. Yuan Jiang, Yale University
- Dr. Ching-Ti Liu, Boston University
- Dr. Xueqin Wang, Sun Yat-Sen University, China
- Dr. Wensheng Zhu, Northeastern Normal University, China

References

- W. Zhu and H. Zhang (2009) Why do we test multiple traits in genetic association studies? (with discussion). *J. Korean Statist. Soc.*, 38:1-10.
- H. Zhang, C.-T. Liu and X. Wang (2010) An association test for multiple traits based on the generalized Kendall's tau. *J. Amer. Statist. Assoc.*, 105:473-481.
- Y. Jiang and H. Zhang (2011) Propensity Score-Based Nonparametric Test Revealing Genetic Variants Underlying Bipolar Disorder. *Genet. Epidemiol.*, 35:125-132.
- H. Zhang (2011) Statistical Analysis in Genetic Studies of Mental Illnesses, *Statistical Science*, in press.
- W. Zhu, Y. Jiang, and H. Zhang (2011) Covariate-Adjusted Association Tests and Power Calculations Based on the Generalized Kendall's Tau. Revised and under review.




Thank You