Genetic Studies of Multivariate Traits

Heping Zhang

Collaborative Center for Statistics in Science
Yale University School of Public Health

Presented at
DIMACS, University of Rutgers

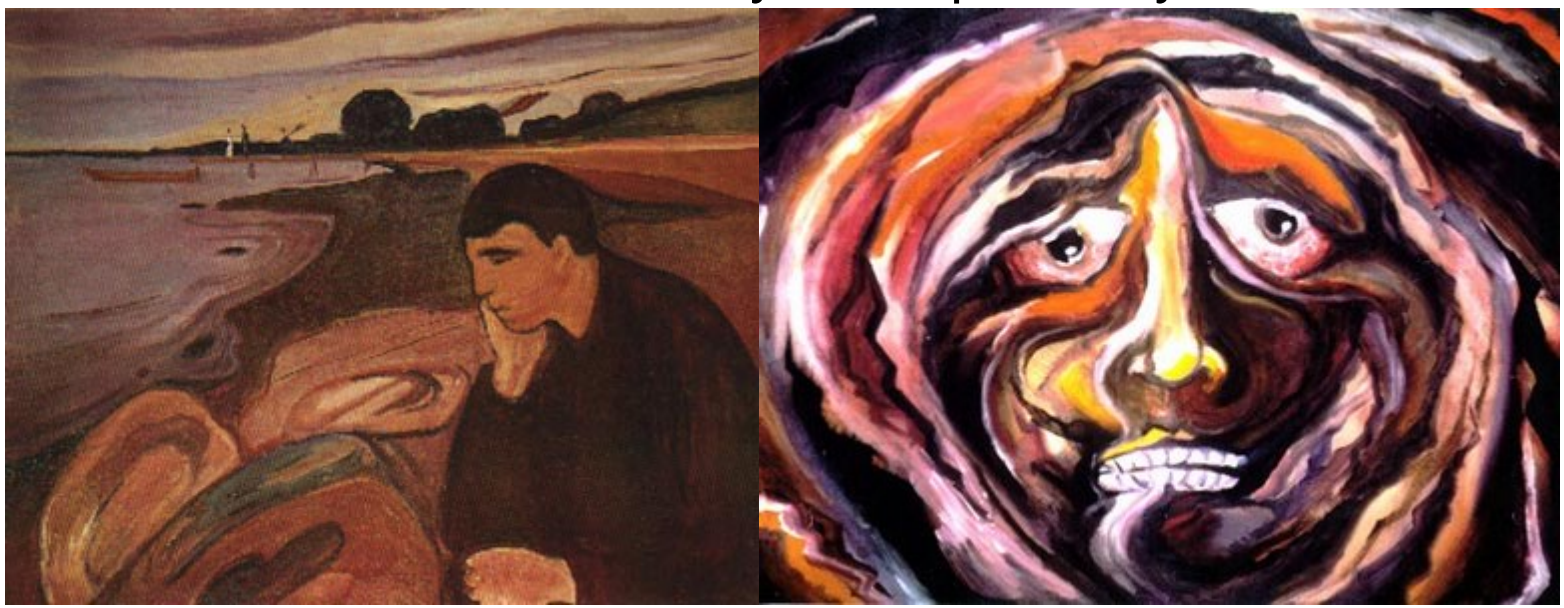
May 17, 2013

Outline

Outline

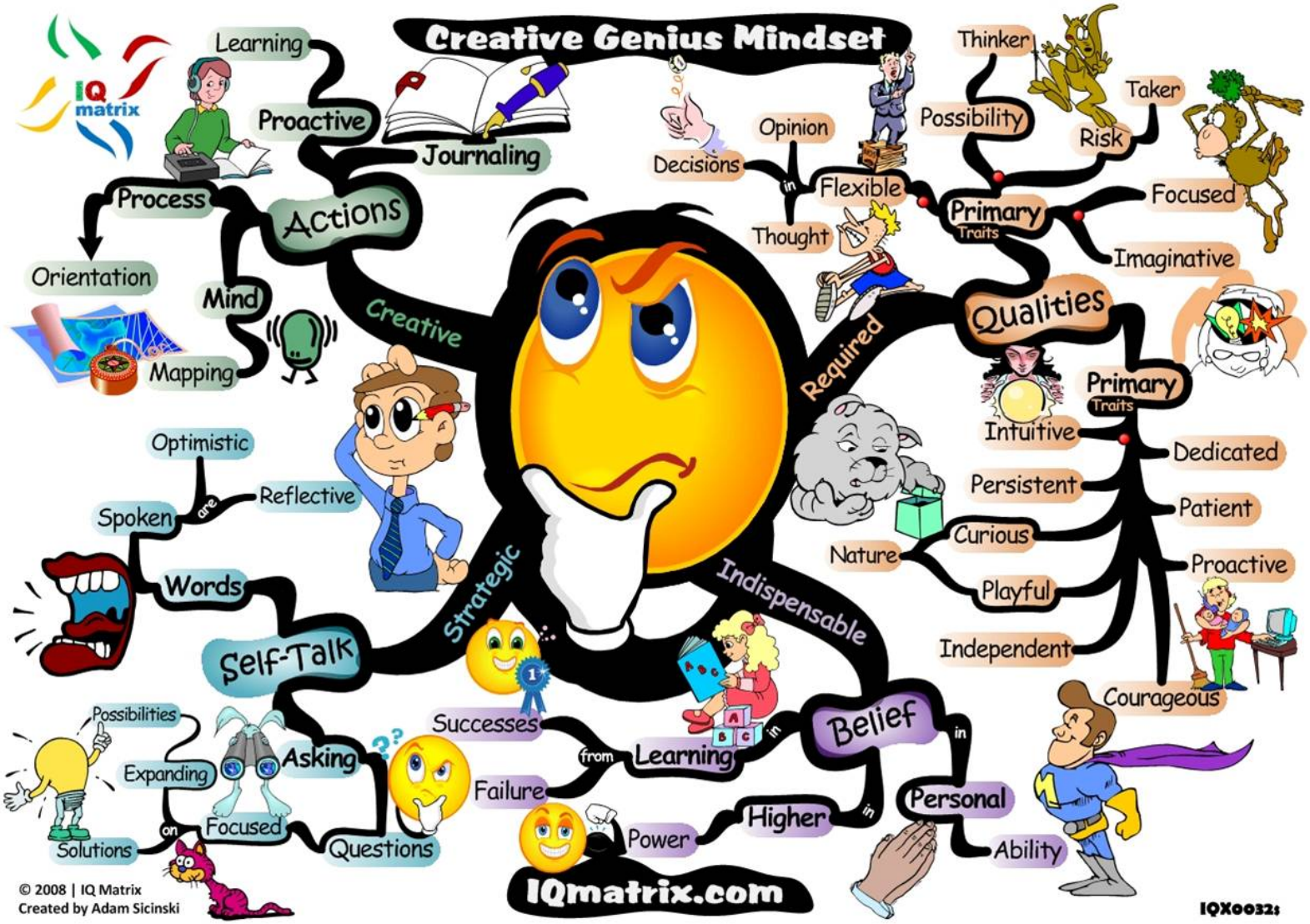
Comorbidity

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

Published 12/2008. Revised 9/2010.

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.

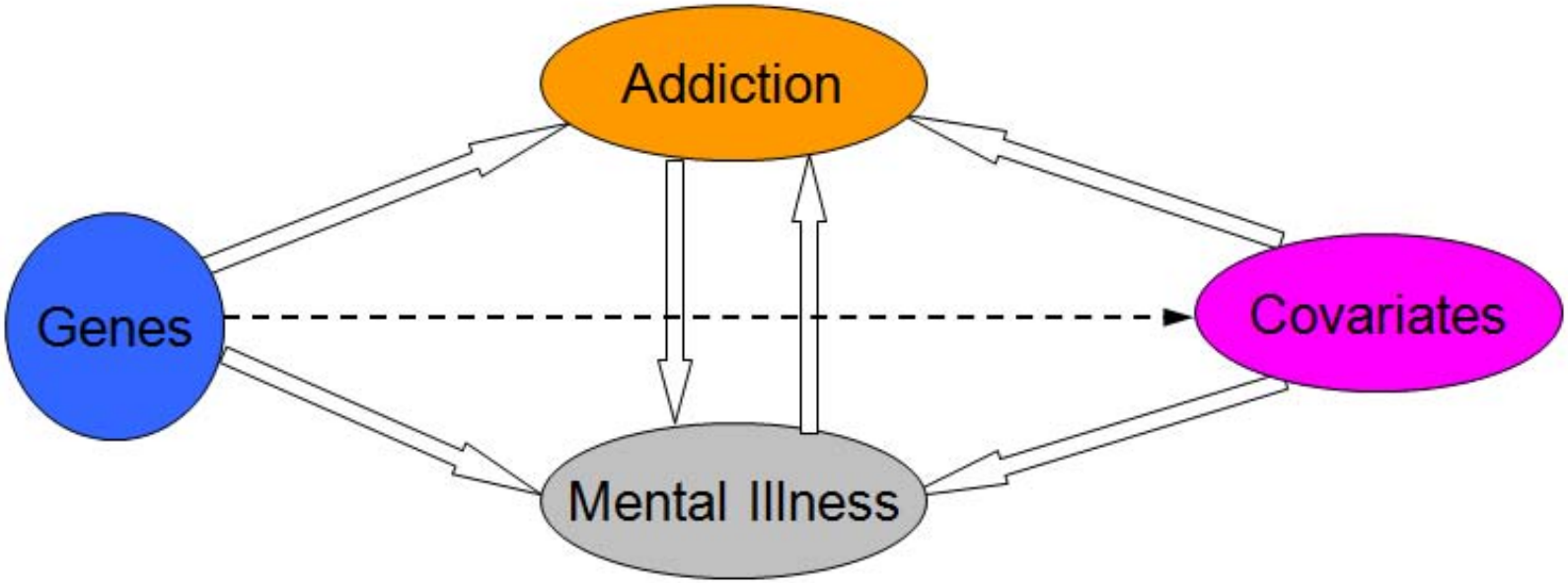
Source: www.nida.nih.gov

Common Etiology for Comorbidity

- Common etiology $\Rightarrow$ $\left\{ \begin{array}{l} \text{mental disorder} \\ \text{drug use disorder} \end{array} \right.$
  - ◇ "Overlapping genetic vulnerabilities: Mounting evidence suggests that common genetic factors may predispose individuals to both mental disorders and addiction or to having a greater risk of the second disorder once the first appears."
  - ◇ "Overlapping environmental triggers: Stress, trauma (e.g., physical or sexual abuse), and early exposure to drugs are common factors that can lead to addiction and to mental illness, particularly in those with underlying genetic vulnerabilities."

Source: www.nida.nih.gov

Disorders, Genes and Covariates



- Covariates: interact or confound genetic effects
- Failure to account for covariates: bias or reduced power
- Null hypothesis: no association between marker alleles and any linked locus that influences traits conditional on covariates

Outline

An Example Trait Questionnaire

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

Ordinal Scale

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

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?

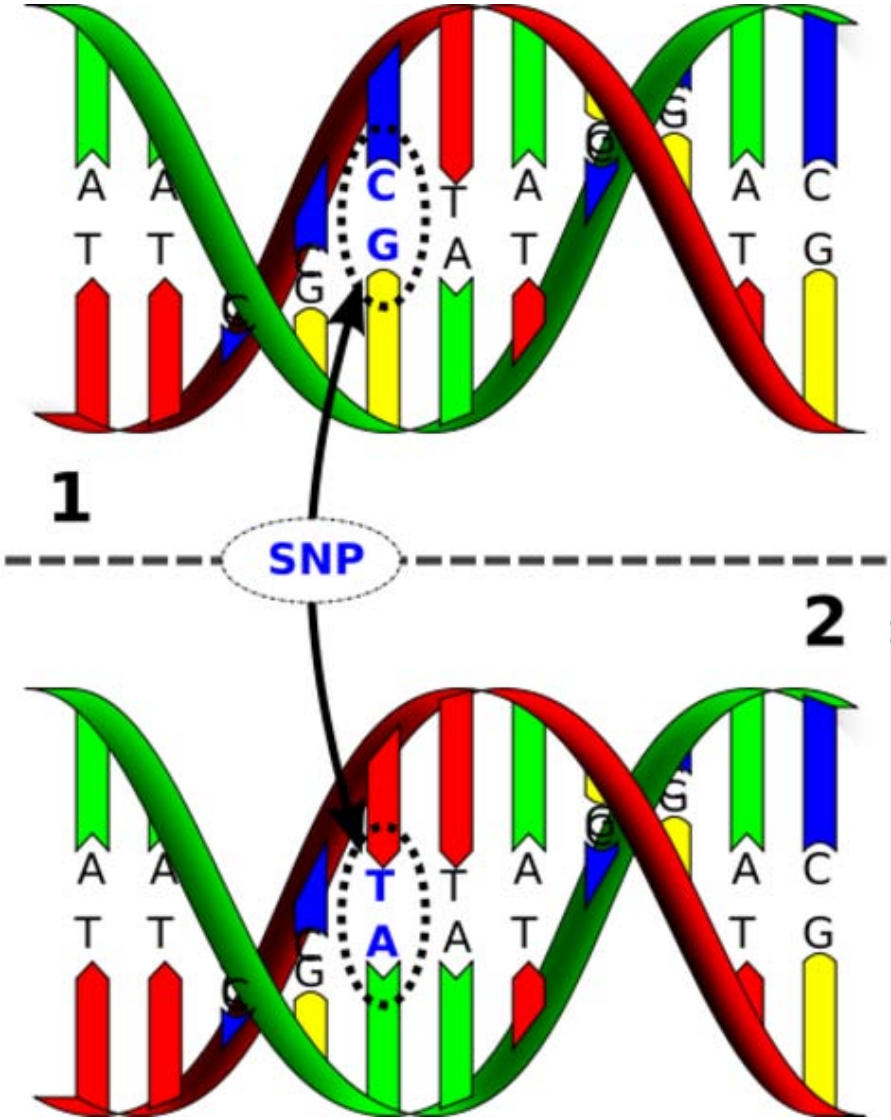
No	**Dichotomous Scale**	0 point	Yes	1 point

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

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.*

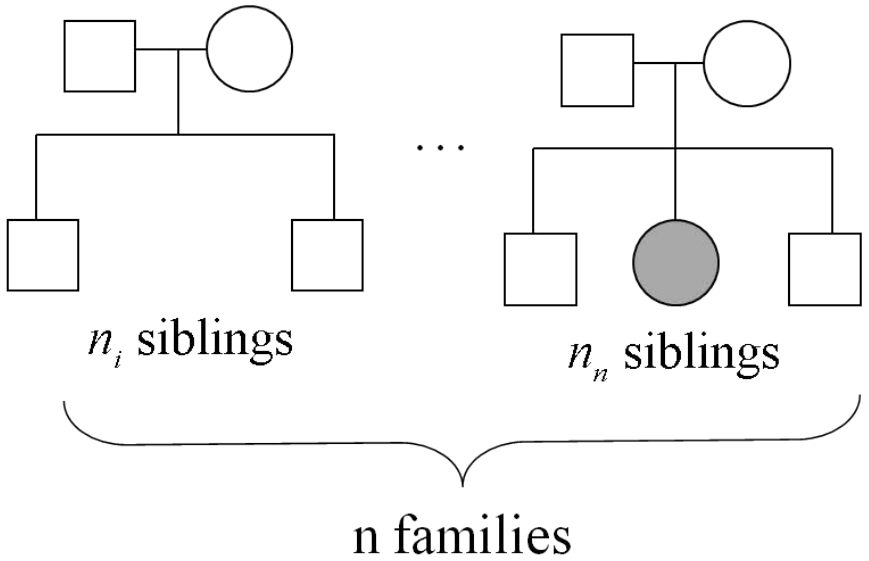
Source: en.wikipedia.org; 2010.census.gov

Study Design

- Population-based studies

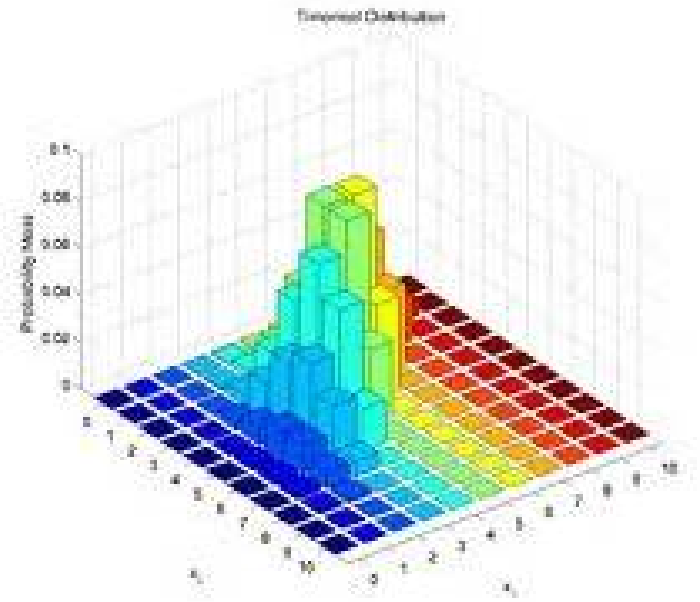
- Family-based studies

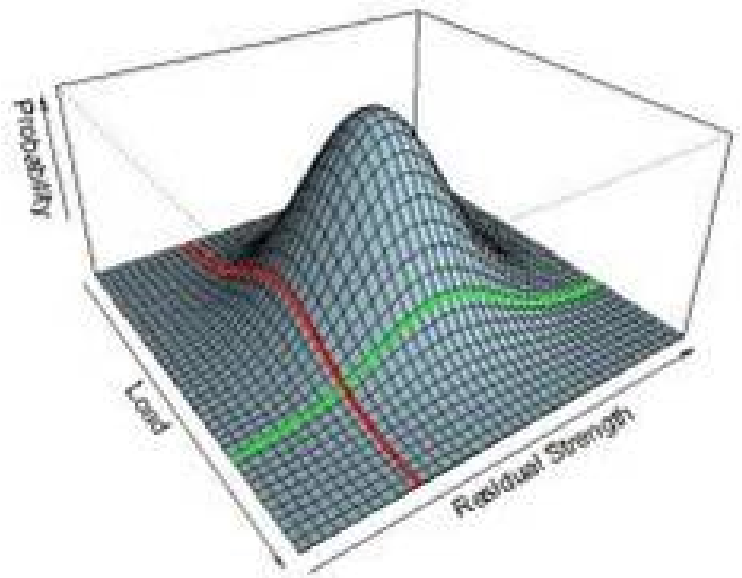
Nuclear families



Outline

Multivariate Distributions





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

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

Challenge

How do we model

a hybrid of
continuous, ordinal, and/or binary responses

???

Outline

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 $\mathbf{T}$, conditional on $\mathbf{Z}$

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)

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: 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

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

An intuitive approach through computation...

- 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., $M^{-1} \sum_{m=1}^{M} I(\tilde{\chi}^2_{\tau,\max,m} \geq \chi^2_{\tau,\max})$

$\Downarrow$ Computation too intensive!

Asymptotic Distribution: Joint Distribution

An theoretical approach through approximation...

- 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* $n \to \infty$*,*

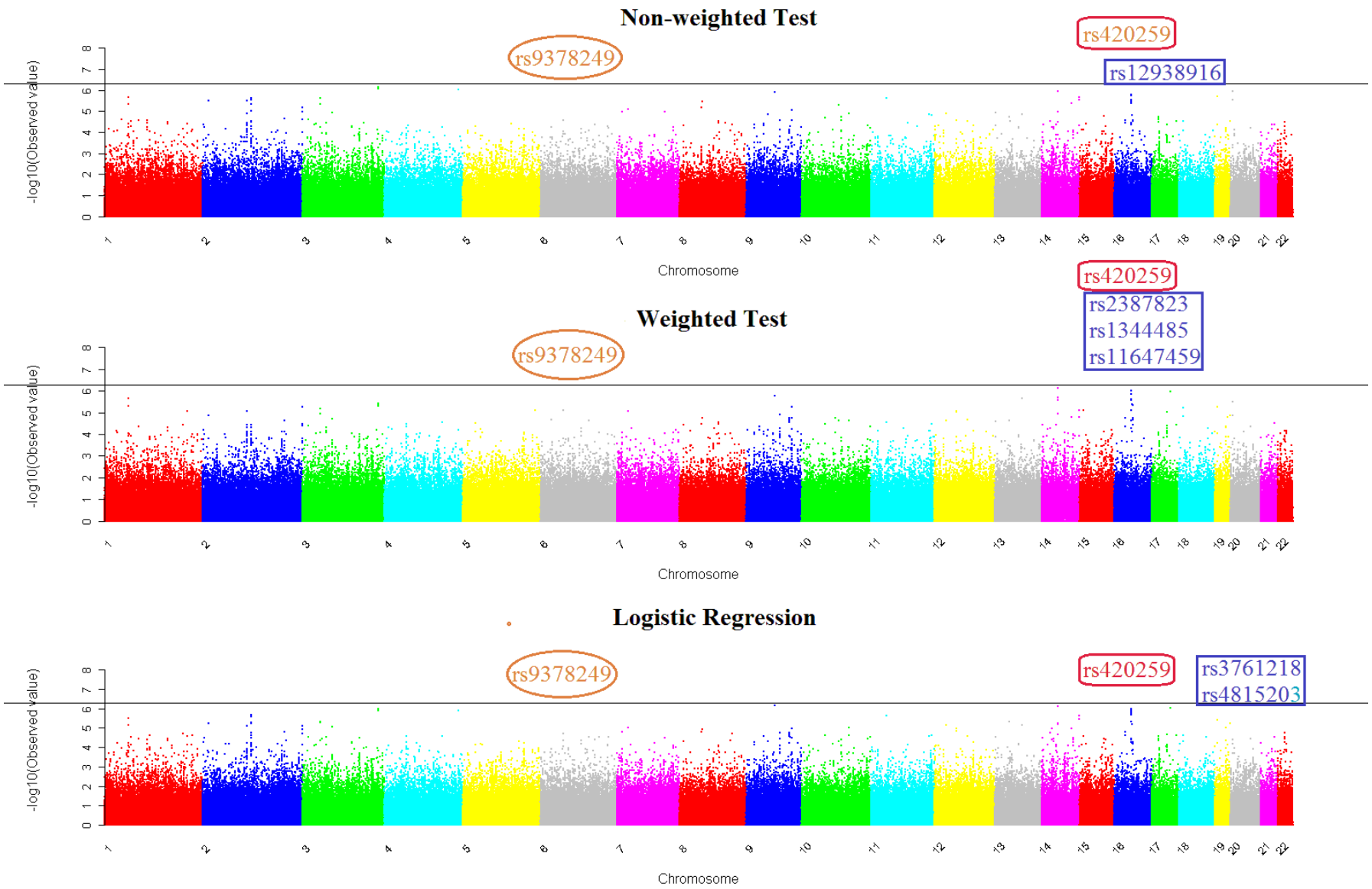
$$\sup_{x\in\mathbb{R}} \left| P\left(\chi^2_{\tau,\max} \leq x\right) - P\left(\max_{1\leq l_1\leq L_1, 1\leq l_2\leq L_2} \|\tilde{\mathbf{R}}_{l_1,l_2}\|^2 \leq x\right)\right| \to 0.$$

Outline

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

Outline

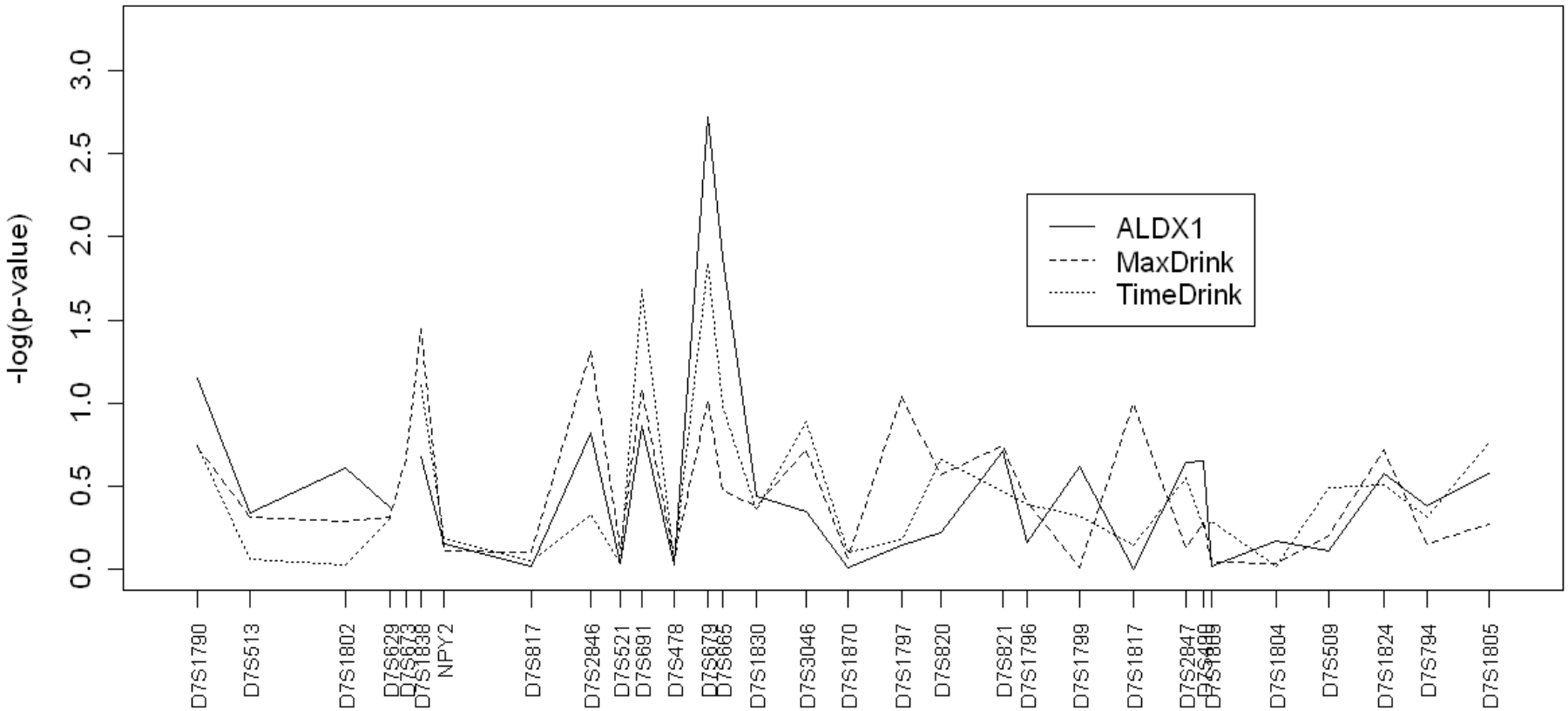
Collaborative Studies on Genetics of Alcoholism

A large scale study to map alcohol dependence susceptible genes

COGA Data

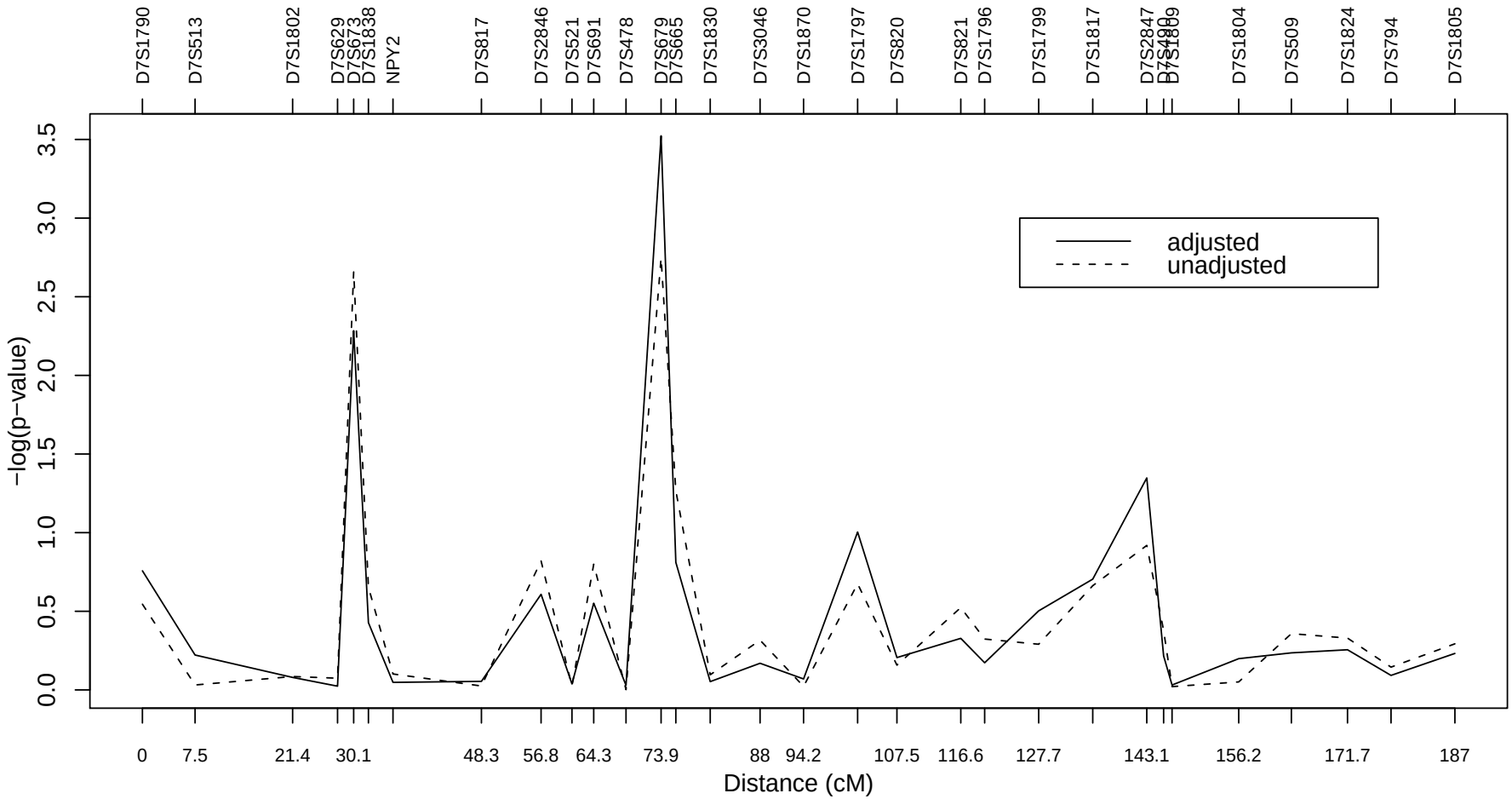
- The data include 143 families with a total of 1,614 individuals
- Multiple Traits:
  - ALDX1 (the severity of the alcohol dependence): pure unaffected, never drunk, unaffected with some symptoms, and affected
  - MaxDrink (maximum number of drinks in a 24 hour period): 0-9, 10-19, 20-29, and more than 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



P-values between one of the three traits and markers on Chromosome 7 with covariates unadjusted.

Results



P-values between the three traits and markers on Chromosome 7 with covariates adjusted or not.

Outline

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 assessing its significance

Outline

Applications

- 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

Outline

Acknowledgment

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

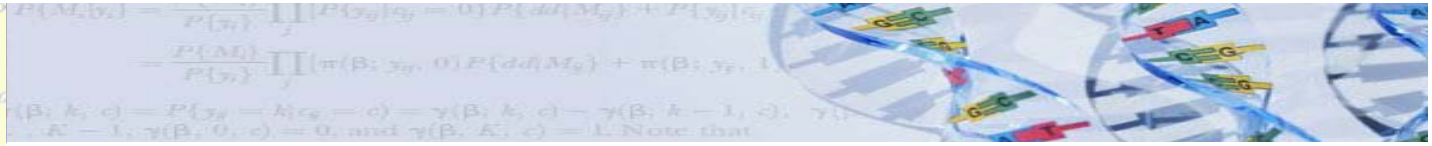
Acknowledgment

- Supported by grant R01DA016750 from National Institute on Drug Abuse.
- The SAGE data were obtained from dbGaP (`http://www.ncbi.nlm.nih.gov`).
- The COGA data were provided by COGA.
- WTCCC data were provided by Wellcome Trust Case-Control Consortium.
- The views expressed here are those of the authors.

Outline

Related Readings

- 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.
- X. Chen, K. Cho, B. Singer, and H. Zhang (2011) The nuclear transcription factor PKNOX2 is a candidate gene for substance dependence in European-origin women. *PLoS One*, 27; 6:e16002.
- 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*, 26: 116-129.
- W. Zhu, Y. Jiang, and H. Zhang (2012) Covariate-Adjusted Association Tests and Power Calculations Based on the Generalized Kendall's Tau. *J. Amer. Statist. Assoc.*, 107:1-11.



"For everything we did, there may be a better way!" – David Banks (?)