Yale

From cross-phenotype associations to pleiotropy in human genetic studies

Andrew DeWan, PhD, MPH
Associate Professor of Epidemiology
Director, Yale Center for Perinatal, Pediatric and Environmental Epidemiology
Yale School of Public Health

Work done in collaboration with Yasmmyn Salinas, PhD, MPH, Assistant Professor of Epidemiology
Yale School of Public Health

Yale SCHOOL OF PUBLIC HEALTH

1

Pleiotropy

- Phenomenon in which a genetic locus affects more than one trait or disease
- Molecular level
  - Single gene with multiple physiological functions
  - Two domains of a single gene product with different functions and affecting multiple phenotypes
  - Gene product with a single function that affects multiple phenotypes acting in multiple tissues
- Statistical level
  - A locus displaying cross-phenotype associations is often considered pleiotropic
  - Can be at the variant, gene or region level

2

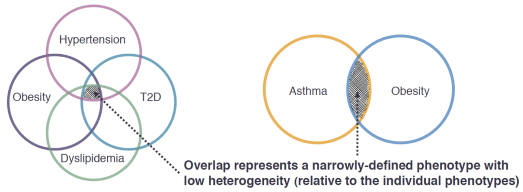
2

Pleiotropy and disease comorbidity

- Examples of correlated (comorbid) disease
  - Obesity, hypertension, dyslipidemia, type 2 diabetes (metabolic disorder)
  - Depression, anxiety, personality disorders (psychiatric disorder)
  - Asthma, obesity (pro-inflammatory conditions)
- Why do certain disease occur together
  - Causality
  - Shared environmental risk factors
  - Shared genetic risk factors

3

Pleiotropy and disease comorbidity



4

Pleiotropy and disease comorbidity

- Pleiotropy-informed analyses consider multiple phenotypes together and take into account the correlation between the phenotypes
  - Analyzing multiple correlated phenotype (e.g. comorbid diseases) is equivalent to analyzing a single narrowly-defined phenotype with low heterogeneity

5

6

Pleiotropy and disease comorbidity

- Detecting shared genetics and/or molecular pathways between comorbid diseases can help us understand exactly how the etiology of the diseases overlap
- Etiologic overlaps:
  - provide opportunities for novel interventions that prevent or treat the comorbidity, rather than preventing/treating each disease separately
  - facilitate drug repurposing (that is, known drugs targeting a pleiotropic locus may be repurposed to treat other diseases controlled by that locus, precluding the need for the development and testing of a brand-new drug)

6

7

Early example of "pleiotropy"

Gregor Mendel documented one of the earliest examples of pleiotropy in his pea plant experiments

Violet flowers
- seed coats = brown-grey
- axils = red and spotted

White flowers
- Seed coats = white
- Axils = white and unspotted

Mendel, J. G., 1866 Experiments in plant hybridization. Verhandlungen des naturforschenden Vereines in Brunn 4: 3–47 (in German).

7

Expression of autism spectrum and schizophrenia in patients with a 22q11.2 deletion

Jacob A.S. Vorstman [a,*], Elemi J. Breetvelt [a], Kirstin I. Thode [b], Eva W.C. Chow [c,d], Anne S. Bassett [c,d]

[a] Rudolf Magnus Institute of Neuroscience, Department of Psychiatry, University Medical Center Utrecht, Utrecht, The Netherlands
[b] Department of Psychiatry, Maiksim Graw Medical Center, Joint Base Andrews, Andrews AFB, MD, USA
[c] Clinical Genetics Research Program, Centre for Addiction and Mental Health, Toronto, Ontario, Canada
[d] Department of Psychiatry, University of Toronto, Toronto, Ontario, Canada

Schizophrenia Research 143 (2013) 55–59

Table 1
Autistic symptoms and probable ASD during childhood, assessed in 77 adults with 22q11.2DS, comparing those with and without schizophrenia.

	Total (n = 77)	22q11.2DS-SZ (n = 36)	22q11.2DS-Co (n = 41)	p				
Mean SRS T-score (95% CI)	73.1 (69.7–76.6)	72.4 (67.3–78.0)	73.7 (68.9–78.6)	0.36[a]				
Subjects categorized as probable ASD	n (%)	n (%)	n (%)	p	Uncorrected OR (95% CI)	p	Corrected OR (95% CI)	p
SRS T score cut-off 60	59 (76.6%)	27 (75.0%)	32 (78.0%)	0.75[b]	0.84 (0.29–2.43)	0.75	0.36 (0.085–1.52)	0.17[c]
SCQ cut-off 15	13 (16.9%)	3 (8.3%)	10 (24.4%)	0.06[b]	0.28 (0.07–1.12)	0.07	0.20 (0.03–1.46)	0.11[d]
SCQ cut-off 12	27 (35.1%)	13 (36.1%)	14 (34.1%)	0.86[b]	1.10 (0.44–2.78)	0.86	1.53 (0.49–5.27)	0.49[d]

CI= confidence interval.
[a] Mann–Whitney–Wilcoxon.
[b] Chi-square.
[c] Corrected for age and IQ.
[d] Corrected for gender, age, IQ.

A B S T R A C T

Background: Copy number variants (CNVs) associated with neuropsychiatric disorders are increasingly being identified. While the initial reports were relatively specific, i.e. implicating vulnerability for a particular neuropsychiatric disorder, subsequent studies suggested that most of these CNVs can increase the risk for more than one neuropsychiatric disorder. Possibly, the different neuropsychiatric phenotypes associated with a single genetic variant are really distinct phenomena, indicating pleiotropy. Alternatively, seemingly different disorders could represent the same phenotype observed at different developmental stages or the same underlying pathogenesis with different phenotypic expressions.
Aims: To examine the relation between autism and schizophrenia in patients sharing the same CNV.
Method: We interviewed parents of 78 adult patients with the 22q11.2 deletion (22q11.2DS) to examine if autistic symptoms during childhood were associated with psychosis in adulthood. We used Chi-square, T-tests and logistic regression while entering cognitive level, gender and age as covariates.
Results: The subgroup of 22q11.2DS patients with probable ASD during childhood did not show an increased risk for psychosis in adulthood. The average SRS scores were highly similar between those with and those without schizophrenia.
Conclusions: ASD and schizophrenia associated with 22q11.2DS should be regarded as two unrelated, distinct phenotypic manifestations, consistent with true neuropsychiatric pleiotropy. 22q11.2DS can serve as a model to examine the mechanisms associated with neuropsychiatric pleiotropy associated with other CNVs.

8

Abundant Pleiotropy in Human Complex Diseases and Traits

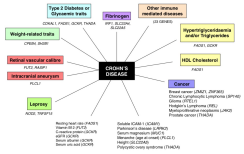
Shanya Sivakumaran,[1,6] Felix Agakov,[1,2,6] Evropi Theodoratou,[1,6] James G. Prendergast,[1] Lina Zgaga,[1,4] Teri Manolio,[5] Igor Rudan,[1] Paul McKeigue,[1] James F. Wilson,[1] and Harry Campbell[1,*]

The American Journal of Human Genetics *89*, 607–618, November 11, 2011

Table 6. Extent of Pleiotropy in Different Disease Classes

	Genes			SNPs		
Disease Class	**Pleiotropic (%)**	**Nonpleiotropic (%)**	**p Value***	**Pleiotropic (%)**	**Nonpleiotropic (%)**	**p Value***
All (comparison group)	233 (16.9)	1147 (83.1)	–	77 (4.6)	1610 (95.4)	–
Immune-mediated phenotypes	106 (37.7)	175 (62.3)	<0.0001	31 (8.3)	343 (91.7)	0.0066
Cancer	49 (34.8)	92 (65.2)	<0.0001	8 (4.8)	158 (95.2)	0.8456
Metabolic syndrome	79 (28.5)	198 (71.5)	<0.0001	30 (8.4)	327 (91.6)	0.0056

* Fisher's exact test p value.



9

10

Pleiotropy in gene mapping

- Mapping a single genotype to multiple phenotypes has the potential to uncover novel links between traits or diseases
- It can also offer insights into the mechanistic underpinnings of known comorbidities
- It can increase power to detect novel associations with one or more phenotypes

10

11

A practitioners' guide for studying pleiotropy in genetic epi studies

Am J Epidemiol. 2017 Aug 11. doi: 10.1093/aje/kwx296. [Epub ahead of print]

Statistical Analysis of Multiple Phenotypes in Genetic Epidemiological Studies:From Cross-Phenotype Associations to Pleiotropy.

Salmas YD, Wang Z, DeWan AT.

Abstract
In the context of genetics, pleiotropy refers to the phenomenon in which a single genetic locus affects more than one trait or disease. Genetic epidemiological studies have identified loci associated with multiple phenotypes, and these cross-phenotype associations are often incorrectly interpreted as examples of pleiotropy. Pleiotropy is only one possible explanation for cross-phenotype associations. Cross-phenotype associations may also arise due to issues related to study design, confounder bias, or non-genetic causal links between the phenotypes under analysis. Therefore, it is necessary to dissect cross-phenotype associations carefully to uncover true pleiotropic loci. In this review, we describe statistical methods that can be used to identify robust statistical evidence of pleiotropy. First, we provide an overview of univariate and multivariate methods for discovery of cross-phenotype associations and highlight important considerations for choosing among available methods. Then, we describe how to dissect cross-phenotype associations by using mediation analysis. Pleiotropic loci provide insights into the mechanistic underpinnings of disease comorbidity, and may serve as novel targets for interventions that simultaneously treat multiple diseases. Discerning between different types of cross-phenotype associations is necessary to realize the public health potential of pleiotropic loci.

© The Author(s) 2017. Published by Oxford University Press on behalf of the Johns Hopkins Bloomberg School of Public Health. All rights reserved. For permissions, please e-mail: journals.permissions@oup.com.

KEYWORDS: genetic epidemiology; mediation analysis; pleiotropy

11

12

Guidelines for generating robust statistical evidence of pleiotropy

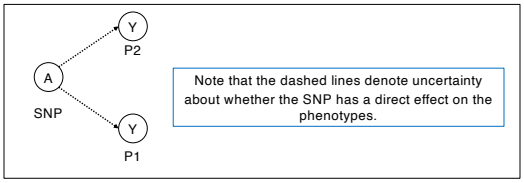
Discover CP associations ⇨ **Dissect** CP associations ⇨ **Classify** them as examples of biological, mediated, or spurious pleiotropy

12

13

Cross-phenotype (CP) associations

Statistical associations between a **single genetic locus** – a single gene or a single variant within a gene – and **multiple phenotypes**

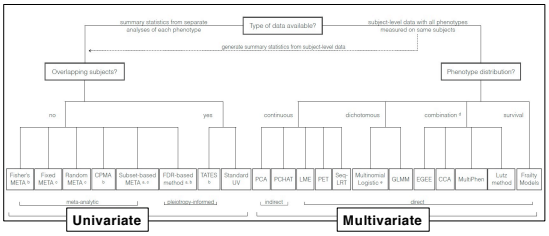


Note that the dashed lines denote uncertainty about whether the SNP has a direct effect on the phenotypes.

13

14

Analytic options for discovery of CP associations



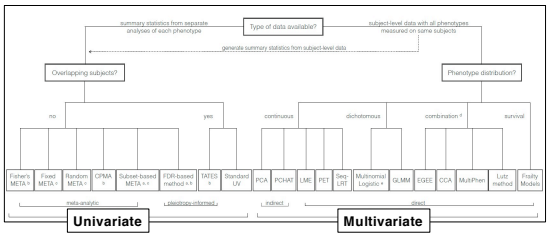
Key distinction:

- Univariate methods examine the association between a given SNP and each trait *separately*
- Multivariate methods examine the association between a given SNP and each trait by modeling the traits *jointly*

14

15

Analytic options for discovery of CP associations



Choice between univariate and multivariate approaches depends on:

- Types of data available on our phenotypes of interest
  - Summary statistics vs. individual-level data?
  - Are the phenotypes measured on the same subjects?
- Distribution of the phenotypes (e.g., quantitative or disease trait)

15

16

Univariate methods are by far the most commonly used to detect CP associations

- Univariate methods include (but are not limited to) the methods you've discussed in class so far:
  - allelic Chi-Square test
  - genotypic Chi-Square test
  - regression-based methods
- The overall approach is to:
  - obtain univariate association p-values for each phenotype
  - declare CP associations at genetic loci that are statistically significantly associated with each phenotype

16

17

Hypothetical example: Discovery of CP associations for hypertension and heart disease by using logistic regression

Step 1. Fit two univariate regression models within PLINK

$$E[hypertension] = \beta_0 + \boldsymbol{\beta_1} * SNP$$
$$E[heart\ disease] = \beta_0 + \boldsymbol{\beta_1} * SNP$$

Word of caution: The univariate tests of association should be marginal tests (conducted irrespectively of the second phenotype) NOT conditional tests (conducted on a subset defined based on absence/presence of the second phenotype). In this example, what that means is that the regression for hypertension should be fit on all subjects *irrespectively* of their heart disease status; and the regression for heart disease should be fit on all subjects *irrespectively* of their hypertension status. More on this later!

evidence to declare a CP association at this SNP.

17

18

Hypothetical example: Discovery of CP associations for hypertension and heart disease by using logistic regression

Step 1. Fit two univariate regression models within PLINK

$$E[hypertension] = \beta_0 + \boldsymbol{\beta_1} * SNP$$
$$E[heart\ disease] = \beta_0 + \boldsymbol{\beta_1} * SNP$$

Step 2. For a given SNP, examine p-values for $\boldsymbol{\beta_1}$ from each model.

- P-value for $\boldsymbol{\beta_1}$ in hypertension model = 1.03×10^{-12}
- P-value for $\boldsymbol{\beta_1}$ in heart disease model = 6.02×10^{-9}

Step 3. Declare CP associations at a given SNP, if the p-values for $\boldsymbol{\beta_1}$ in each model surpass the study significance threshold.

- Assuming the standard GWAS significance threshold (alpha=5×10^{-8}), there is a statistically significant association with both hypertension and heart disease at this particular SNP. Therefore, we have sufficient statistical evidence to declare a CP association at this SNP.

18

Using multivariate methods to increase the power to detect cross-phenotype associations

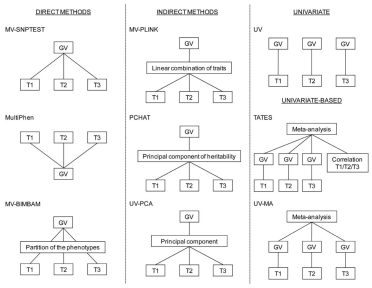
19

A Comparison of Multivariate Genome-Wide Association Methods

Tessel E. Galesloot[1], Kristel van Steen[2,3], Lambertus A. L. M. Kiemeney[1,4], Luc L. Janss[5*], Sita H. Vermeulen[1,4*]

1 Department for Health Evidence, Radboud university medical center, Nijmegen, The Netherlands, 2 Systems and Modeling Unit, Montefiore Institute, University of Liège, Liège, Belgium, 3 Bioinformatics and Modeling, GIGA-R, University of Liège, Liège, Belgium, 4 Department of Urology, Radboud university medical center, Nijmegen, The Netherlands, 5 Department of Molecular Biology and Genetics, Aarhus University, Aarhus, Denmark, 6 Department of Human Genetics, Radboud university medical center,

PLOS ONE | www.plosone.org 1 April 2014 | Volume 9 | Issue 4 | e95923

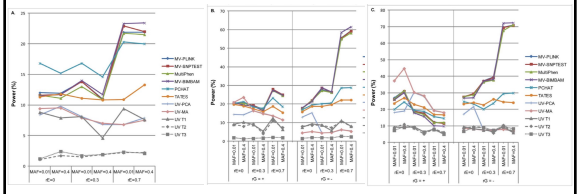


20

Table 1. Simulation scenarios.

# traits associated with QTL	Heritability (h^2_j)	Effect size (a_j)	rG	rE	MAF (q)
1	$h^2_1 = 0.1\%$, $h^2_2 = h^2_3 = 0$	$a_1 > 0$, $a_2 = a_3 = 0$	0	$3 \times 0/3 \times 0.3/3 \times 0.7$	0.01/0.4
2	$h^2_1 = h^2_2 = 0.1\%$, $h^2_3 = 0$	$a_1 = a_2$, $a_3 = 0$	+	$3 \times 0/3 \times 0.3/3 \times 0.7$	0.01/0.4
	$h^2_1 = h^2_2 = 0.1\%$, $h^2_3 = 0$	$-a_1 = a_2$, $a_3 = 0$	−	$3 \times 0/3 \times 0.3/3 \times 0.7$	0.01/0.4
3	$h^2_1 = h^2_2 = h^2_3 = 0.1\%$	$a_1 = a_2 = a_3$	+	$3 \times 0/3 \times 0.3/3 \times 0.7$	0.01/0.4
	$h^2_1 = h^2_2 = h^2_3 = 0.1\%$	$-a_1 = a_2 = a_3$	−	$3 \times 0/3 \times 0.3/3 \times 0.7$	0.01/0.4

MAF indicates minor allele frequency; j, trait; QTL, quantitative trait locus; rE, residual correlation; rG, genetic correlation.
doi:10.1371/journal.pone.0095923.t001



21

A comparison of univariate and multivariate GWAS methods for analysis of multiple dichotomous phenotypes

Yasmmyn D. Salinas[1], Andrew T. DeWan[1], and Zuoheng Wang[2]

[1] Department of Chronic Disease Epidemiology; [2] Department of Biostatistics, Yale School of Public Health, Yale University, 60 College St, New Haven, Connecticut, USA

Genet. Epidemiol. 41 (7), 689-689

22

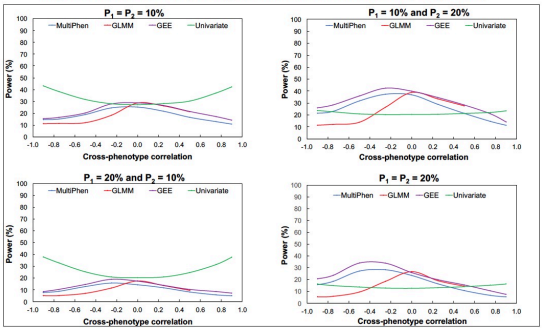
Simulation scenarios

# traits associated	h_j^2	$r_{Y1,Y2}$	P_j
1	h_1^2=0.1%,h_2^2=0%	[-0.9,0.9]	P1 = P2 = 10%
			P1 = P2 = 20%
			P1 = 10%, P2 = 20%
			P1 = 20%, P2 = 10%
2	$h_1^2 = h_2^2$= 0.1%	[-0.9,0.9]	P1 = P2 = 10%
			P1 = P2 = 20%
			P1 = 10%, P2 = 20%
			P1 = 20%, P2 = 10%
2	h_1^2 = 0.1%,h_2^2 = 0.05%	[-0.9,0.9]	P1 = P2 = 10%
			P1 = P2 = 20%
			P1 = 10%, P2 = 20%
			P1 = 20%, P2 = 10%

23

PLEIOTROPY PRESENT
equal effect sizes

Figure 2. Power when both phenotypes are associated with the SNP ($h_1^2 = h_2^2 = 0.1\%$) [a]



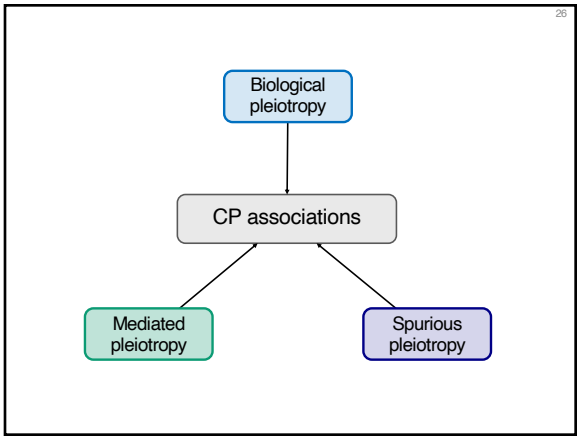
[a] Results for GLMMs are shown for $r_{Y1,Y2} \leq 0.5$ only, since the models experienced convergence issues for $r_{Y1,Y2} > 0.5$.

24

25

Problem: CP associations need not be indicative of pleiotropy

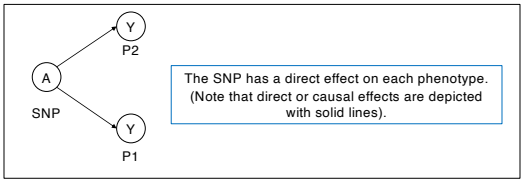
25



26

Biological pleiotropy

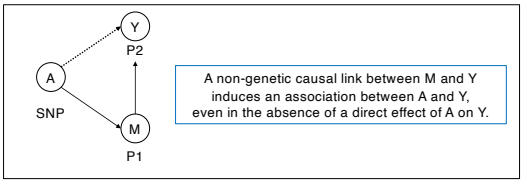
Independent associations between a genetic locus (A) and multiple phenotypic outcomes (Y)



27

Mediated pleiotropy

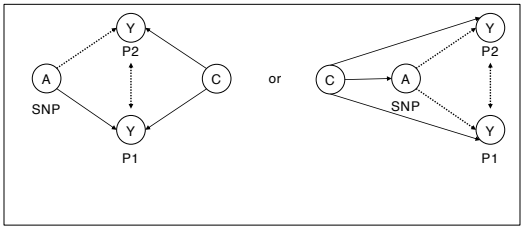
Association between a genetic locus (A) and an intermediate phenotype (M) that causes a second phenotypic outcome (Y)



28

Spurious pleiotropy

Artifactual associations with multiple phenotypes due to issues related to study design, confounding, or associations with markers in strong linkage disequilibrium* with multiple causal variants in different genes

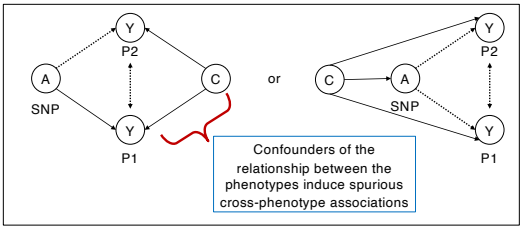


*Linkage disequilibrium is the non-random co-segregation of alleles.

29

Spurious pleiotropy

Artifactual associations with multiple phenotypes due to issues related to study design, confounding, or associations with markers in strong linkage disequilibrium* with multiple causal variants in different genes



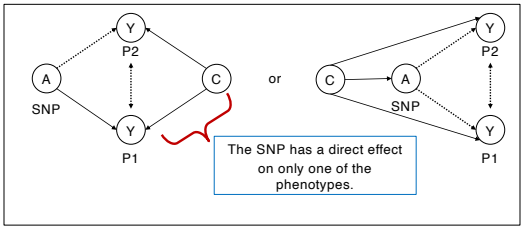
*Linkage disequilibrium is the non-random co-segregation of alleles.

30

31

Spurious pleiotropy

Artifactual associations with multiple phenotypes due to issues related to study design, confounding, or associations with markers in strong linkage disequilibrium* with multiple causal variants in different genes



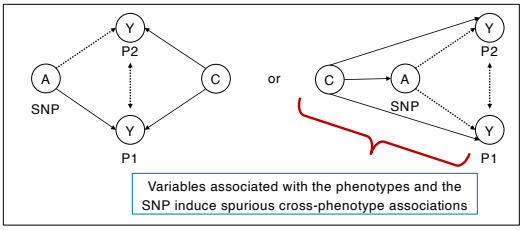
*Linkage disequilibrium is the non-random co-segregation of alleles.

31

32

Spurious pleiotropy

Artifactual associations with multiple phenotypes due to issues related to study design, confounding, or associations with markers in strong linkage disequilibrium* with multiple causal variants in different genes



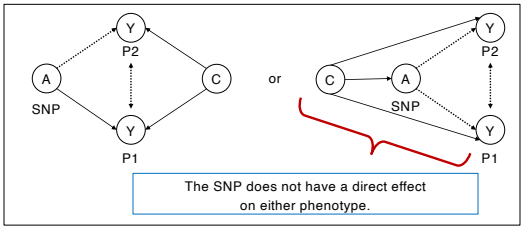
*Linkage disequilibrium is the non-random co-segregation of alleles.

32

33

Spurious pleiotropy

Artifactual associations with multiple phenotypes due to issues related to study design, confounding, or associations with markers in strong linkage disequilibrium* with multiple causal variants in different genes



*Linkage disequilibrium is the non-random co-segregation of alleles.

33

34

Guidelines for generating robust statistical evidence of pleiotropy

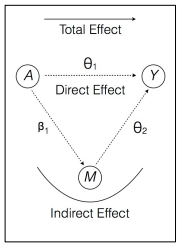
Discover CP associations → **Dissect** CP associations → Classify them as examples of biological, mediated, or spurious pleiotropy

34

35

Mediation analysis provides a tool for dissecting CP associations

- Mediation analysis decomposes the **total effect** of the SNP (*A*) on a phenotypic outcome (*Y*) into:
  - **Direct effect:** effect of *A* on *Y* that occurs independently of an intermediate phenotype (*M*)
  - **Indirect effect:** effect of *A* on *Y* that occurs through the intermediate phenotype *M*

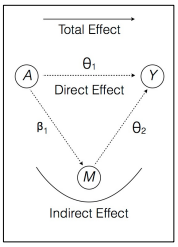


35

36

Mediation analysis: Data requirements

- All phenotypes must be measured on the same subjects
- Temporality must be ascertained
  - The occurrence of the intermediate variable *M* must precede that of the phenotypic outcome variable *Y*

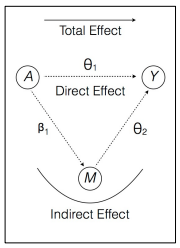


36

37

Mediation analysis: Assumptions

- There must be no unmeasured:
  - confounders of the total effect
  - confounders of the relationship between SNP *A* and the mediator *M*
  - confounders of the relationship between mediator *M* and phenotypic outcome *Y*



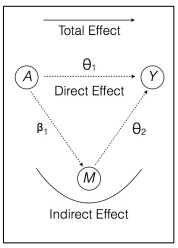
37

38

Mediation analysis: Assumptions

Typically met in genetic epi studies!

- There must be no unmeasured:
  - confounders of the total effect
  - confounders of the relationship between SNP *A* and the mediator *M*
  - confounders of the relationship between mediator *M* and phenotypic outcome *Y*

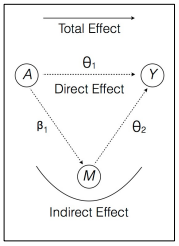


38

39

Mediation analysis: Assumptions

- There must be no unmeasured:
  - confounders of the total effect
  - confounders of the relationship between SNP *A* and the mediator *M*
  - confounders of the relationship between mediator *M* and phenotypic outcome *Y*



Requires adjustment for known confounders to prevent bias
(Note: this effectively restricts the use of mediation analyses to datasets in which data on such variables have been collected)

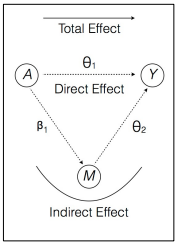
39

40

Mediation analysis: Regression-based approach

- Requires fitting two regression models, one for mediator *M* and one for phenotypic outcome *Y*:
  - $E[M\,|a,c] = \beta_0 + \boldsymbol{\beta_1} a + \beta_2' c$
  - $E[Y\,|a,m,c\,] = \theta_0 + \boldsymbol{\theta_1} a + \boldsymbol{\theta_2} m + \theta_4' c$

Assesses the effect of *A* on *M*, while controlling for measured confounders *(C)*



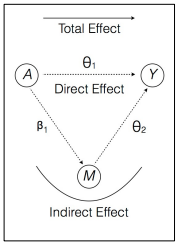
40

41

Mediation analysis: Regression-based approach

- Requires fitting two regression models, one for mediator *M* and one for phenotypic outcome *Y*:
  - $E[M\,|a,c] = \beta_0 + \boldsymbol{\beta_1} a + \beta_2' c$
  - $E[Y\,|a,m,c\,] = \theta_0 + \boldsymbol{\theta_1} a + \boldsymbol{\theta_2} m + \theta_4' c$

Assesses the effect of *A* on *Y*, while controlling for both *M* and *C*

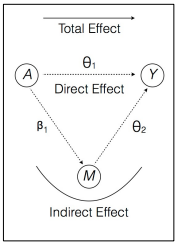


41

42

Mediation analysis: Regression-based approach

- Requires fitting two regression models, one for mediator *M* and one for phenotypic outcome *Y*:
  - $E[M\,|a,c] = \beta_0 + \boldsymbol{\beta_1} a + \beta_2' c$
  - $E[Y\,|a,m,c\,] = \theta_0 + \boldsymbol{\theta_1} a + \boldsymbol{\theta_2} m + \theta_4' c$
- The parameter estimates from these models (**namely $\boldsymbol{\beta_1}$, $\boldsymbol{\theta_1}$, and $\boldsymbol{\theta_2}$**) are used to estimate the direct and indirect effects



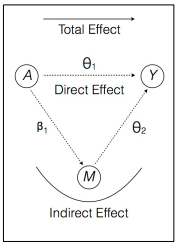
42

Guidelines for generating robust statistical evidence of pleiotropy

43

Mediation analysis: Interpretation

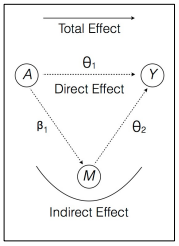
- **Biological pleiotropy:** SNP A is associated with mediator M, and the total effect of SNP A on phenotypic outcome Y is equal to its direct effect (i.e., the indirect effect is equal to 0)



44

Mediation analysis: Interpretation

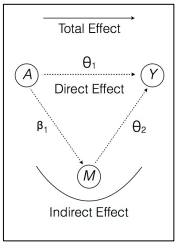
- **Mediated pleiotropy**
  - Complete mediation:
    - SNP *A* is associated with mediator *M* and the total effect of *A* on phenotypic outcome *Y* is equal to its indirect effect (i.e., the direct effect is equal to 0).
  - Incomplete mediation:
    - SNP *A* is associated with mediator *M* and *A* has both direct and indirect effects on phenotypic outcome *Y* (i.e., the total effect is equal to the sum of the direct and indirect effects)



45

Mediation analysis: Interpretation

- **Spurious pleiotropy**
  - SNP A is not associated with mediator M after controlling for measured confounders



46

mediation R package

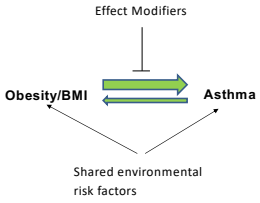
```
> med.fit<-glm(W1~rs1_2, data=combined, family=binomial("logit"))
> out.fit<-glm(W2~W1+rs1_2, data=combined, family=binomial("logit"))
> med.out<-mediate(med.fit,out.fit, treat="rs1_2", mediator="W1", boot=TRUE, boot.ci.type="bca", sims=1000)
> summary(med.out)

Causal Mediation Analysis

Nonparametric Bootstrap Confidence Intervals with the BCa Method

                    Estimate 95% CI Lower 95% CI Upper p-value
ACME (control)      0.02152  0.01823     0.03  <2e-16 ***
ACME (treated)      0.02199  0.01868     0.03  <2e-16 ***
ADE (control)       0.00723  0.00415     0.01  <2e-16 ***
ADE (treated)       0.00771  0.00443     0.01  <2e-16 ***
Total Effect        0.02922  0.02461     0.03  <2e-16 ***
Prop. Mediated (control)  0.73634  0.65429     0.84  <2e-16 ***
Prop. Mediated (treated)  0.75247  0.67272     0.85  <2e-16 ***
ACME (average)      0.02175  0.01847     0.03  <2e-16 ***
ADE (average)       0.00747  0.00426     0.01  <2e-16 ***
Prop. Mediated (average)  0.74441  0.66254     0.84  <2e-16 ***
```

47

Empirical searches for pleiotropic loci for asthma and obesity

48

Asthma-obesity comorbidity



Ford ES. The epidemiology of obesity and asthma. J Allergy Clin Immunol. 2005;115(5):897-909; quiz 10.
Stukus DR. Obesity and asthma: The chicken or the egg? J Allergy Clin Immunol. 2016.
Kim SH, Sutherland ER, Gelfand EW. Is there a link between obesity and asthma? Allergy Asthma Immunol Res. 2014;6(3):189-95.
Egan KB, Ettinger AS, DeWan AT, Holford TR, Holmen TL, Bracken MB. Longitudinal associations between asthma and general and abdominal weight status among Norwegian adolescents and young adults: the HUNT Study. Pediatric obesity. 2016.

49

Am J Hum Genet. 2009 Jul;85(1):87-96. doi: 10.1016/j.ajhg.2009.06.011. Epub 2009 Jul 2.

PRKCA: a positional candidate gene for body mass index and asthma.

Murphy A[1], Tantisira KG, Soto-Quirós ME, Avila L, Klanderman BJ, Lake S, Weiss ST, Celedón JC.

Study design

- Two phases:
  - genome-wide linkage analysis of BMI
  - follow-up family-based candidate-gene association study of BMI and asthma
- Strategy for candidate-gene study:
  - Authors focused on a single gene (*PRKCA*) within the BMI linkage peak because:
    - animal models suggest role of PRKCA in obesity; and
    - published association studies of other genes within the linkage peak had found no association with BMI.

50

Study population

- Costa Rica study
  - N = 415 asthmatic children + parents
- Childhood Asthma Management Program
  - N = 493 non-Hispanic White asthmatic children + parents

Note that ALL children in both study populations are asthmatic

51

Phenotype definitions

- Body mass index (BMI)
  - calculated from objective measures of height and weight
- Asthma
  - physician-diagnosed asthma + one of the following:
    - 2 respiratory symptoms or asthma attacks in prior year
    - increased airway responsiveness or bronchodilator response

52

Statistical methods

- Univariate family-based association tests (FBATs) were used to test *PRKCA* SNPs for association with BMI and asthma *separately*
  - Note: The FBAT statistic takes into account the phenotype of the **offspring only**
- Significance threshold used by study authors: $\alpha = 9.5 \times 10^{-5}$

53

Results for BMI

Table 3. Evidence for Association of *PRKCA* with BMI in Costa Rica and CAMP

Marker	Location (BP)[a]	Minor Allele	Allele Frequency: CR	Allele Frequency: CAMP	Number of Informative Families[b] (number of offspring with 0/1 recoded genotype): CR	Number of Informative Families[b] (number of offspring with 0/1 recoded genotype): CAMP	Effect Size[c]: CR	Effect Size[c]: CAMP	CR p Value[d,e]	CAMP Replication p Value[d,e] (two-sided)	Joint p Value[f] (CR, CAMP two-sided)
rs228883	61874457	T	0.27	0.33	91 (67/24)	110 (80/39)	2.45	1.60	+0.0011	+0.0038 (+0.0076)	$5.6 \times 10^{-5**}$ (1.0×10^{-4})
rs1005651	61868473	C	0.26	0.33	83 (60/23)	113 (83/39)	2.27	1.60	+0.0019	+0.0039 (+0.0077)	$9.5 \times 10^{-5**}$ (1.8×10^{-4})
rs228875	61924337	A	0.29	0.35	101 (70/31)	129 (92/46)	1.71	1.22	+0.0109	+0.0182 (+0.0364)	0.0019 (0.0035)
rs2244497	61931405	C	0.31	0.36	120 (86/34)	136 (98/47)	1.69	1.21	+0.0160	+0.0171 (+0.0341)	0.0025 (0.0046)

Two BMI-associated variants

54

Results for asthma

Table 4. Evidence for Association of *PRKCA* with Asthma in Costa Rica and CAMP

Marker	Location (BP)[a]	Minor Allele	Allele Frequency: CR	Allele Frequency: CAMP	Number of Informative Families[b] (number of offspring with 0/1 recoded genotype): CR	Number of Informative Families[b] (number of offspring with 0/1 recoded genotype): CAMP	Costa Rica p Value[c,d]	CAMP Replication p Value[c,d] (two-sided)	Joint p Value[e] (CR, CAMP two-sided)
rs732191	61779673	G	0.46	0.35	168 (117/51)	141 113/43	−0.0194	−0.0214 (−0.0428)	0.0036 (0.0067)
rs9895580	61789701	C	0.47	0.35	168 (117/51)	141 114/43	−0.0171	−0.0160 (−0.0320)	0.0025 (0.0047)
rs4411531	61793662	A	0.29	0.12	88 (70/18)	25 (24/1)	−0.0058	−0.0058 (−0.0117)	0.0004 (0.0007)
rs8080771	61824330	G	0.46	0.35	164 (116/48)	108 (90/29)	−0.0161	−0.0070 (−0.0140)	0.0011 (0.0021)
rs11652956	61839798	G	0.29	0.12	83 (65/18)	23 (22/1)	−0.0101	−0.0111 (−0.0222)	0.0011 (0.0021)
rs7221968	61848731	C	0.27	0.11	79 (63/16)	18 (17/1)	−0.0122	−0.0216 (−0.0432)	0.0024 (0.0045)
rs7405806	61862056	A	0.49	0.31	164 (109/55)	90 (77/20)	−0.0309	−0.0009 (−0.0018)	0.0003 (0.0006)
rs11079657	61862528	A	0.38	0.23	129 (94/35)	60 (56/8)	−0.0092	−0.0002 (−0.0004)	2.6×10^{-5}** (5.0×10^{-5}**)

One asthma-associated variant

55

Conclusions

- **Authors' conclusion: *PRKCA* displays pleiotropy for asthma and BMI (pleiotropy at gene level)**
  - Two variants (rs228883 and rs1005651) displayed statistically significant associations with body mass index
  - A different variant (rs11079657) displayed a statistically significant association with asthma.

56

Conclusions

- **Our conclusion: *PRKCA* is associated with asthma and with BMI among asthmatics (no true CP association!)**
  - There is insufficient evidence to declare a CP association at *PRKCA* because the test of association with BMI was not a marginal test
    - FBAT test for BMI only took into account the phenotype of the offspring – which were ALL asthmatic
  - Thus, it remains to be seen whether the association with BMI is also present among non-asthmatics subjects
  - Without that information, we would not be able to assess whether asthma is a **mediator** or a **moderator** of the relationship between *PRKCA* and BMI.

57

A GWAS study of pleiotropy

Discovery and Mediation Analysis of Cross-Phenotype Associations Between Asthma and Body Mass Index in 12q13.2

Yasmmyn D. Salinas*, Zuoheng Wang, and Andrew T. DeWan

* Correspondence to Dr. Yasmmyn D. Salinas, Department of Chronic Disease Epidemiology, Yale School of Public Health, 60 College Street, New Haven, CT 06520 (e-mail: yasmmyn.salinas@yale.edu).

Am J Epidemiol. 2021;190(1):85–94

58

Study design

- Two parts:
  - Genome-wide search for cross-phenotype associations with asthma and body mass index
  - Follow-up mediation analysis to dissect genome-wide significant CP associations

59

Study population

- N = 305,945 White, British subjects from the UK Biobank (a population-based prospective cohort study of > 500,000 subjects, aged 40-69 years at baseline)

biobank uk
Improving the health of future generations

60

61

Phenotype definitions

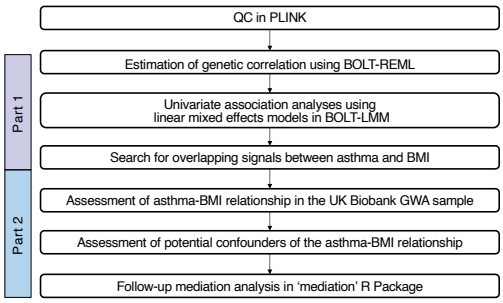
- BMI at baseline (kg/m^2):
  - calculated based on height and weight measurements collected by trained UK Biobank staff at the recruitment sites
- Asthma diagnosed prior to baseline (yes/no):
  - ascertained via the question "Has a doctor ever told you that you had asthma?"
  - Note: In mediation analyses, two subgroups were created based on age-at-diagnosis

biobank uk
Improving the health of future generations

61

62

Statistical Methods



62

63

Overlap in GWA signals

Association with BMI among the 1,457 SNPs with genome-wide significant p-values for asthma

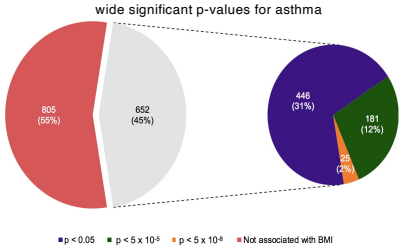


Figure 1. Overlap in GWA signals between asthma and BMI. Results for asthma are for the analysis of all asthmatic subjects (35,373 asthmatics vs. 270,572 non-asthmatics). Results for BMI are for the quantitative BMI analysis (n=305,945). Both analyses are sex- and age-adjusted. The threshold for genome-wide significance was alpha=5x10^{-8}.

63

64

Overlap in GWA signals

Association with asthma among the 1,699 SNPs with genome-wide significant p-values for BMI

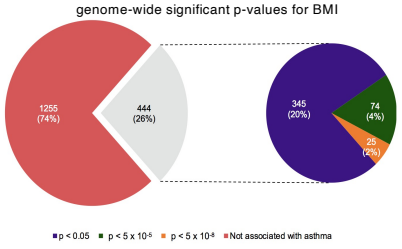
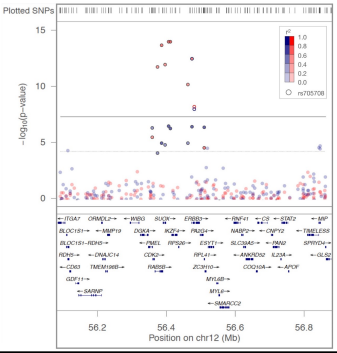


Figure 1. Overlap in GWA signals between asthma and BMI. Results for asthma are for the analysis of all asthmatic subjects (35,373 asthmatics vs. 270,572 non-asthmatics). Results for BMI are for the quantitative BMI analysis (n=305,945). Both analyses are sex- and age-adjusted. The threshold for genome-wide significance was alpha=5x10^{-8}.

64

Regional plot around rs705708 for BMI (blue) and asthma (red)



65

65

66

Cross-phenotype associations in 12q13.2

Table 2. Cross-phenotype associations in 12q13.2 [a]

					Asthma		BMI[c]	
SNP	Gene	BP	Effect/reference allele	EAF	OR (95% CI)	P[c]	beta (95% CI)	P[d]
rs2069408	*CDK2*	56,364,321	G/A	0.3388	1.04 (1.02, 1.06)	3.30×10^{-6}	-0.06 (-0.08, -0.04)	5.40×10^{-7}
rs1873914	*RAB5*	56,379,427	C/G	0.4237	1.06 (1.04, 1.08)	2.40×10^{-12}	-0.05 (-0.07, -0.02)	7.90×10^{-5}
rs705702 [b]	*SUOX*	56,390,636	G/A	0.3376	1.07 (1.05, 1.09)	3.10×10^{-14}	-0.05 (-0.08, -0.03)	1.10×10^{-5}
rs10876864 [a]	*SUOX*	56,401,085	G/A	0.4279	1.06 (1.04, 1.08)	1.50×10^{-12}	-0.05 (-0.07, -0.03)	1.60×10^{-5}
rs1701704	*IKZF4*	56,412,487	G/T	0.3433	1.07 (1.05, 1.09)	1.50×10^{-14}	-0.06 (-0.09, -0.04)	3.70×10^{-7}
rs2456973	*IKZF4*	56,416,928	C/A	0.3432	1.07 (1.05, 1.09)	1.50×10^{-14}	-0.06 (-0.08, -0.04)	6.00×10^{-7}
rs11171739 [b]	*ERBB3*	56,470,625	C/T	0.4337	1.06 (1.04, 1.07)	8.80×10^{-11}	-0.05 (-0.07, -0.03)	1.10×10^{-5}
rs2292239	*ERBB3*	56,482,180	T/G	0.3470	1.07 (1.05, 1.08)	4.50×10^{-13}	-0.06 (-0.08, -0.04)	4.20×10^{-7}
rs705708	*ERBB3*	56,488,913	A/G	0.4712	1.05 (1.03, 1.07)	7.20×10^{-9}	-0.06 (-0.09, -0.04)	1.30×10^{-6}
rs11171747 [b]	*ESYT1*	56518408	T/G	0.6180	1.04 (1.02, 1.05)	2.90×10^{-6}	-0.06 (-0.08, -0.04)	4.50×10^{-7}

Abbreviations: BP = base-pair ; BMI = body mass index; CI = confidence interval; EAF = effect allele frequency; OR = odds ratio; SNP = single-nucleotide polymorphism

a. Results shown for SNPs with p < 5×10^{-8} for asthma and p < 0.05 for BMI.
b. For intergenic SNPs, the nearest gene is listed, with priority given to genes directly downstream of variant.
c. P-value from BOLT-LMM, derived using the standard "infinitesimal" mixed model.
d. P-value from BOLT-LMM, derived using the Gaussian mixture model.

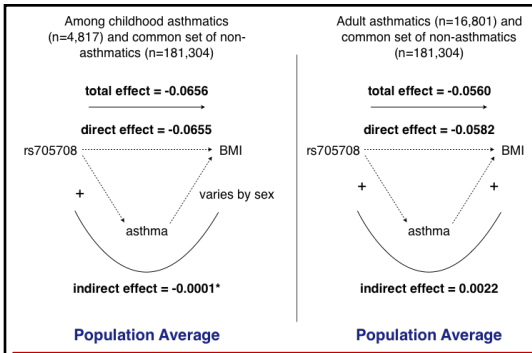
66

67

Decomposing the effect of rs705708 on BMI via mediation analysis

67

68



Note: Effect estimates shown are adjusted for common determinants of asthma and BMI: age, sex, breast-feeding status, exposure to maternal smoking, and smoking status at asthma diagnosis (adult analyses only). Unless otherwise noted by an asterisk(*), all paths are significant at the 0.05 level.

68

69

Conclusions

- rs705708 has a positive direct effect on asthma
  - Stronger in magnitude for childhood asthma
- rs705708 has a negative direct effect on BMI
  - Consistent in magnitude and direction in analyses including childhood vs. adult asthmatics
- This suggests that locus 12q13.2, tagged by rs705708, has pleiotropic effects on asthma and BMI.

69

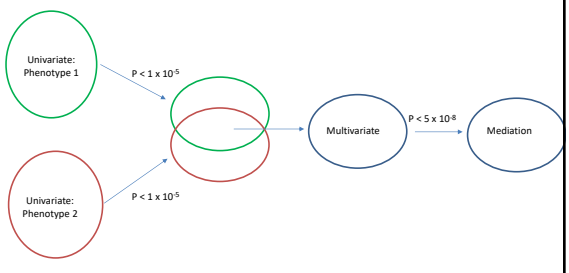
70

Conclusions

- 12q13.2 is multigenic and our CP associations span genes *CDK2, RAB5*, *SUOX*, *IZK4*, *RPS26, ERBB3*, and *ESYT1*.
  - rs705708 is the top regional BMI signal and resides in *ERBB3*.
  - The top regional asthma signal, rs2456973, resides in *IZKF4*.
  - While rs705708 and rs2456973 could be in LD with the same causative variant in either *ERBB3* or *IKZF4* or another gene in 12q13.2, it is also possible that each variant could tag a distinct, trait-specific causative variant in different genes.
- Therefore, locus 12q13.2 displays pleiotropic effects on asthma and BMI, but this may not be an example of pleiotropy at the gene level (biological pleiotropy).

70

71



71