Measuring the Phenotype:
What disease endpoint or trait are you studying?

Erin M. Ramos, PhD MPH
Office of Population Genomics
NHGRI

July 18, 2008

Course and Lecture Objectives

- **Course Objective #2:** *To understand the various methods, their advantages, and their disadvantages for the definition of phenotypes (disease endpoints, quantitative traits, etc) for use in association studies.*

- **Learning Objectives:**
  - Convey the importance of selecting appropriate phenotypes for your genomic research study
  - Describe the properties of a good measure and the consequences of measurement error on study results
  - Consider the advantages of using standard measures for your phenotype of interest

Lecture Outline

1) **Phenotype definition**
   1. Discrete verse quantitative traits
   2. Complex disease and natural history of disease
   3. Selecting your phenotype
2) **Measurement error**
   1. Properties of a good measure
   2. Consequences of measurement error
3) **Advantages of using standard phenotypes**
   1. Why is it important to use standard measures?
   2. Example of successful cross-study analyses
   3. Introduction to PhenX

1. Phenotype *(φαινότυπος)*

- Means “the form of what appears”
- Root *φαίνειν* (phanein) also found in *φαινόμενον* (phenomenon)
- Also linked to *φως*, *φωτός* (light, of the light)
- In order for something to appear, we need light to see it
- A phenotype is the observable expression of an individual’s genotype

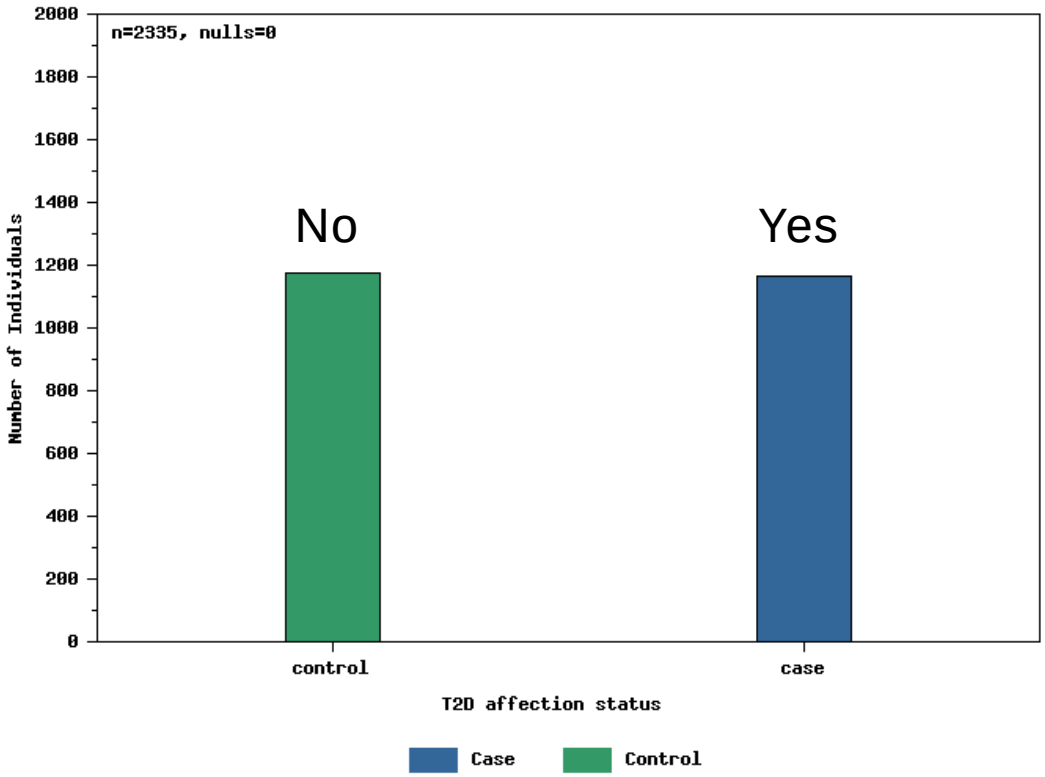
‘In writing the history of a disease…[T]he clear and natural phenomenon of the disease should be noted … accurately, and in all their minuteness; in imitation of the exquisite industry of those painters who represent in their portraits the smallest moles and faintest spots.’

-T. Sydenham (Medical Observations, 3rd ed. London, 1749)

Discrete Trait

- **Discrete/Dichotomous**
  - Two values
  - e.g. Type II Diabetes (No/Yes)
  - Typically of direct clinical relevance (e.g. cancer, hypertension, arthritis)

Distribution of Measured Values for Type II Diabetes



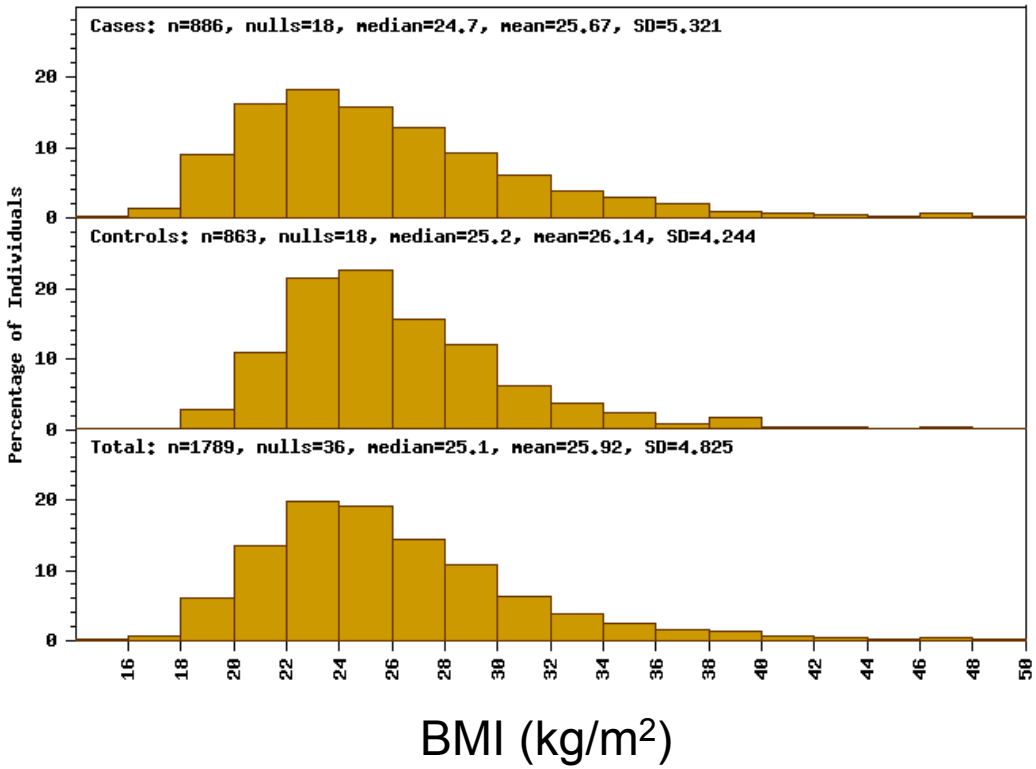
dbGaP: http://www.ncbi.nlm.nih.gov/sites/entrez?db=gap

Data from The Finland-United States Investigation of NIDDM Genetics (Fusion) Study

Quantitative Trait

- **Quantitative/Continuous**
  - Range of possible values (e.g. Systolic blood pressure, BMI)
  - Can be reduced to a discrete/dichotomous trait by using a predefined threshold value

Distribution of Measured Values for BMI

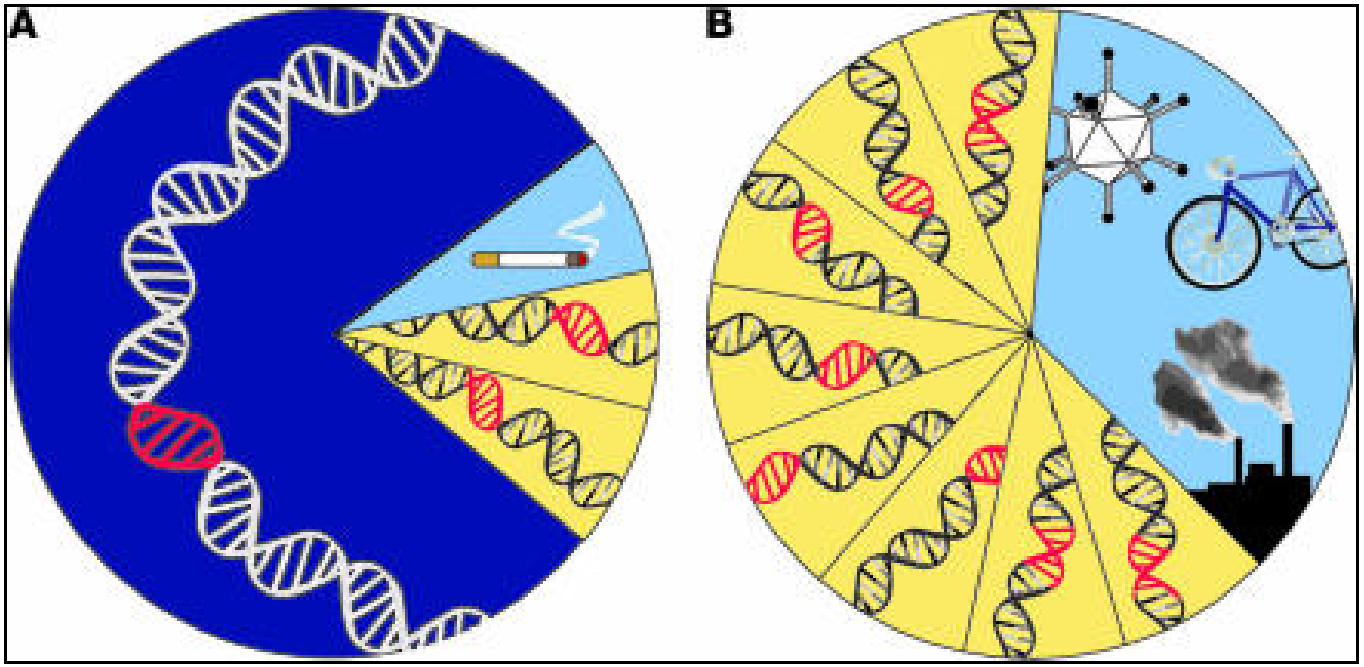


BMI (kg/m^2)

dbGaP: http://www.ncbi.nlm.nih.gov/sites/entrez?db=gap

Data from the GAIN: Search for Susceptibility Genes for Diabetic Nephropathy in Type 1 Diabetes Study

Gene & Environment Contribution to Disease



A. Monogenic Disease. A variant in a single gene is the primary determinant of a monogenic disease or trait, responsible for most of the disease risk or trait variation (dark blue sector), with possible minor contributions of modifier genes (yellow sectors) or environment (light blue sector). **B. Complex disease.** Many variants of small effect (yellow sectors) contribute to disease risk or trait variation, along with many environmental factors (blue sector).

(Manolio TA, et al, JCI, 118: 1590, 2008)

Complex Disease

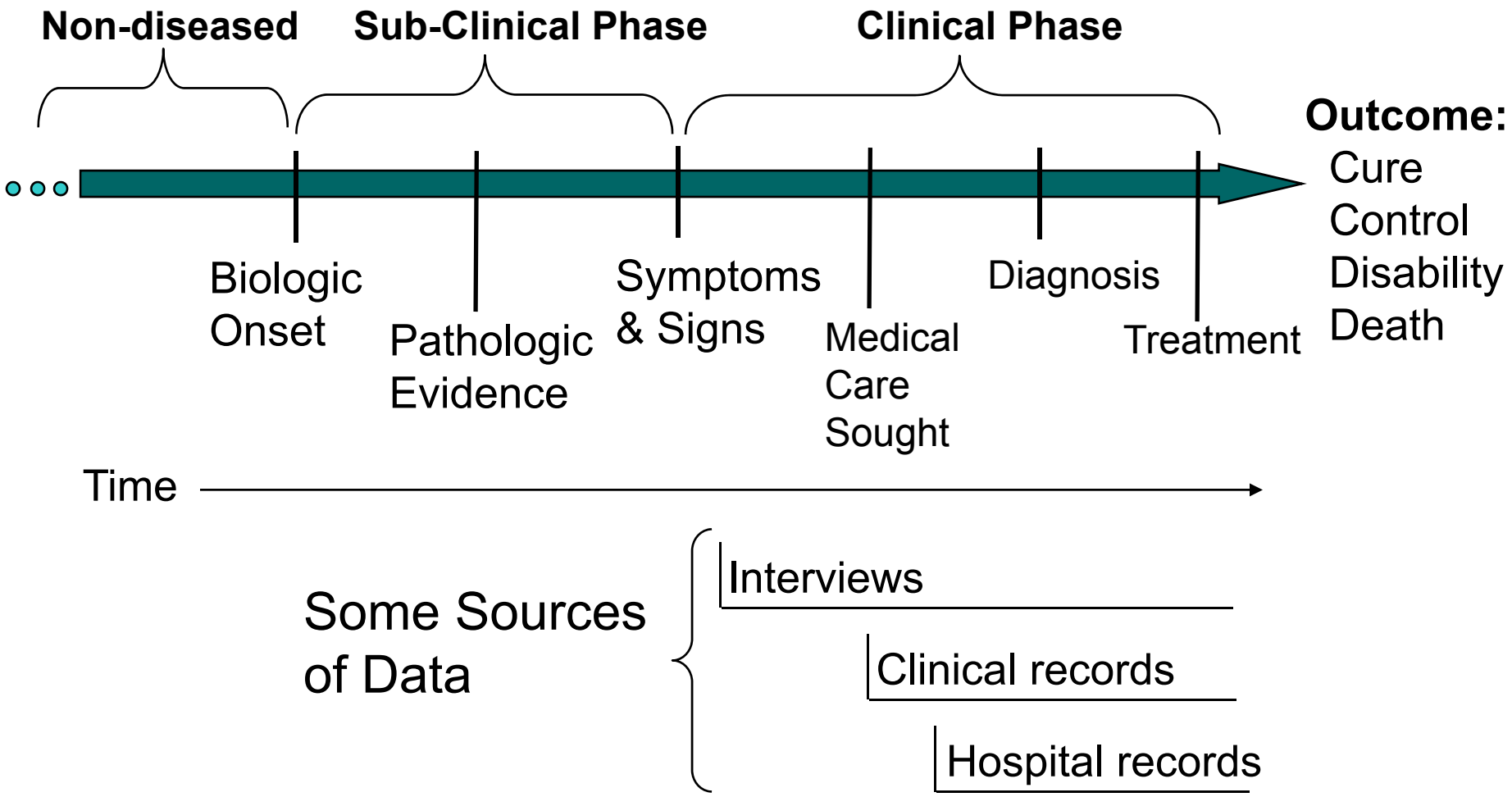
- Characterized by high levels of genetic complexity; multiple genes may act independently or interact with other genes to influence the phenotype
- Multiple manifestations with varying degrees of genetic influence
  - e.g. Myocardial Infarction, Coronary Artery Atherosclerosis, and Sudden Cardiac Death are forms of Coronary Disease
- Multiple causes, which may be attributed to separate or overlapping genetic influences
  - e.g. Atherosclerosis caused by lipid accumulation, inflammation, endothelial disruption, and thrombosis.

Complex Disease

- Difficult to distinguish individuals with “sub-clinical” disease from “non-diseased” individuals if early stage diagnosis is inadequate
- Characterized by variable age of onset of clinical symptoms
- Environmental factors may modify the genotype-phenotype relationship; thus, disease expression range from nearly undetectable to severely debilitating

(Ellsworth DL and Manolio TA, Annals of Epi, 1999)

Natural history of disease



(Adapted from Gordis, 3rd Ed, 2004)

Some Limitations of Hospital Data

- Hospital admissions are selective in relation to:
  - Demographics
  - Severity of disease
  - Associated conditions
  - Admission patterns
- Hospital records are not typically designed for research. They may be:
  - Incomplete, illegible, or missing
  - Variable in diagnostic quality
- Populations at risk are not generally defined

Some Limitations of Clinical Data

- Can be a rich source of patient specific data (clinical exam, diagnostic tests, and procedures), but…
- Chart extraction can be difficult
- Patients might receive care from additional sources
- Uneven organization, incompleteness, legibility, etc…
- Clinical diagnostic criteria can vary and change over time

Some Limitations of Interview Data

- The respondent:
  - Has the disease, but does not have symptoms and does not report the disease
  - Has the disease, sought medication attention, but reports a different disease
  - May provide disease information accurately, but it is recorded inaccurately
- The interviewer may know the hypothesis being tested, thus probing more intensively in one group of respondents than another
- Incomplete or missing data

Selecting your phenotype

Goal: Reduce heterogeneity in your phenotype to increase your chance of finding genes!!

- What disease/trait interests you?
- Evidence for genetic influence on your disease/trait of interest
- Homogenous cases (highly specific disease criteria)
- Intermediate phenotypes (closer proximity to genes)

Evidence for genetic influence

Familial Clustering:

- Risk of disease in relative of case > risk in relative of non-case or general population
- Discrete Trait: Familial relative risk, Risch's λ_s
- Continuous Trait: parent-offspring correlation & sib-sib correlation

- Twin studies
  - Comparing Concordance between Monozygotic Twins and Dizygotic Twins

Association of rs10033464 & Atrial Fibrillation (AF)

Variants conferring risk of atrial fibrillation on chromosome 4q25

Daniel F. Gudbjartsson[1], David O. Arnar[2], Anna Helgadottir[1], Solveig Gretarsdottir[1], Hilma Holm[2], Asgeir Sigurdsson[1], Adalbjorg Jonasdottir[1], Adam Baker[1], Gudmar Thorleifsson[1], Kristleifur Kristjansson[1], Arnar Palsson[1], Thorarinn Blondal[1], Patrick Sulem[1], Valgerdur M. Backman[1], Gudmundur A. Hardarson[1], Ebba Palsdottir[1], Agnar Helgason[1], Runa Sigurjonsdottir[2], Jon T. Sverrisson[3], Konstantinos Kostulas[4], Maggie C. Y. Ng[5], Larry Baum[5], Wing Yee So[5], Ka Sing Wong[5], Juliana C. N. Chan[5], Karen L. Furie[6], Steven M. Greenberg[6], Michelle Sale[6], Peter Kelly[6], Calum A. MacRae[7], Eric E. Smith[6], Jonathan Rosand[6], Jan Hillert[4], Ronald C. W. Ma[5], Patrick T. Ellinor[7], Gudmundur Thorgeirsson[2], Jeffrey R. Gulcher[1], Augustine Kong[1], Unnur Thorsteinsdottir[1] & Kari Stefansson[1]

- Discovery Study (Iceland Cases):
  - All cases of AF at two large hospitals
- Replication Study (U.S. Cases):
  - Younger patients with lone AF
  - AF with co-existing hypertension
  - Stroke patients with AF

(Gudbjartsson DF et al, Nature, June 2007)

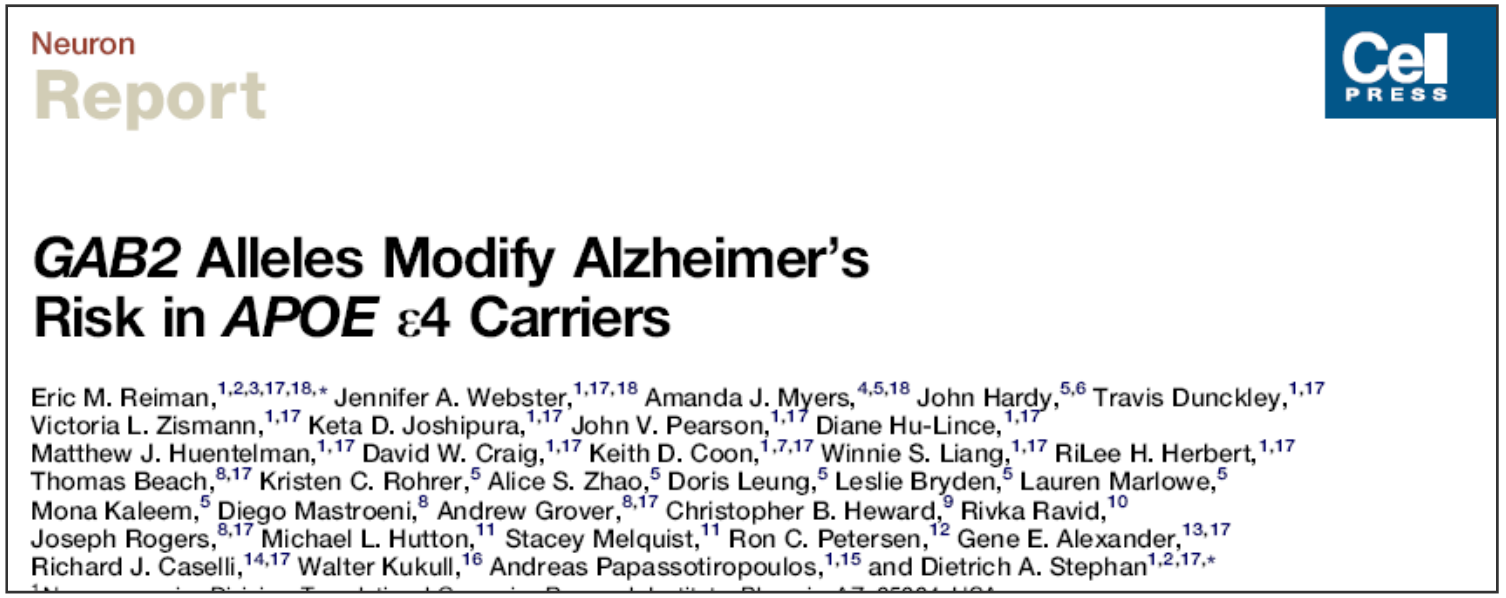
Association of rs10033464* & Atrial Fibrillation (AF)

	Case / Control	Mean Age (yr)	OR	*p*-value
Iceland	2801 / 17,714		1.40	9.4×10^{-9}
United States				
Lone AF	251 / 804	46.1	1.68	1.2×10^{-10}
AF & Hyp	67 / 804	54.5	1.66	.001
Other AF	318 / 804	75.2	0.97	.015

OR = Odds Ratio; **PITX2* gene, known to be involved in early heart development

(Gudbjartsson DF et al, Nature, June 2007)

Association of GAB2 alleles & LOAD



Neuron
Report

Cell PRESS

GAB2 Alleles Modify Alzheimer's Risk in *APOE* ε4 Carriers

Eric M. Reiman,[1,2,3,17,18,*] Jennifer A. Webster,[1,17,18] Amanda J. Myers,[4,5,18] John Hardy,[5,6] Travis Dunckley,[1,17] Victoria L. Zismann,[1,17] Keta D. Joshipura,[1,17] John V. Pearson,[1,17] Diane Hu-Lince,[1,17] Matthew J. Huentelman,[1,17] David W. Craig,[1,17] Keith D. Coon,[1,7,17] Winnie S. Liang,[1,17] RiLee H. Herbert,[1,17] Thomas Beach,[8,17] Kristen C. Rohrer,[5] Alice S. Zhao,[5] Doris Leung,[5] Leslie Bryden,[5] Lauren Marlowe,[5] Mona Kaleem,[5] Diego Mastroeni,[8] Andrew Grover,[8,17] Christopher B. Heward,[9] Rivka Ravid,[10] Joseph Rogers,[8,17] Michael L. Hutton,[11] Stacey Melquist,[11] Ron C. Petersen,[12] Gene E. Alexander,[13,17] Richard J. Caselli,[14,17] Walter Kukull,[16] Andreas Papassotiropoulos,[1,15] and Dietrich A. Stephan[1,2,17,*]

- Late-Onset Alzheimer's Disease (LOAD)
  - Discovery set: clinically & neuropathologically confirmed LOAD cases and neuropathologically normal controls
  - Rationale: to exclude misdiagnosed cases & cognitively normal controls who have LOAD neuropathology

(Reiman et al, *Neuron* 2007; 54:713-720)

Association of GAB2 alleles & LOAD

Stage	Cases (N)	Controls (N)	SNPs
1	446 Neuropathology Discovery	290	312,316
2	197 Neuropathology Replication	114	"
3	218 Clinical Replication*	146	"
Total	861	550	7

* To confirm genetic association independent of brain donor selection bias

(Reiman et al, *Neuron* 2007; 54:713-720)

Association of GAB2 alleles & LOAD

SNP	P-Value	Freq. in Controls	OR	[95%CI
rs1385600	2.81 E-09	0.71	3.65	[2.34,5.71]
rs1007837	3.97 E-07	0.73	3.01	[1.94,4.68]
rs4945261	3.08 E-08	0.72	3.44	[2.18,5.43]
rs10793294	1.59 E-07	0.66	2.83	[1.90,4.21]
rs4291702	5.88 E-07	0.70	2.96	[1.91,4.59]
rs7115850	2.80 E-10	0.67	3.92	[2.51,6.11]
rs2373115	9.66 E-11	0.70	4.06	[2.81,14.69]

* Sample Size = 861 Cases & 550 Controls

(Reiman et al, *Neuron* 2007; 54:713-720)

Rs2373115 interacts w/ APOE to modify risk

*APOE*e4* Group	*APOE*e4* OR [95% CI]	rs2373115 OR [95%CI]
*APOE*e4* -		1.12 [0.82,1.53]
*APOE*e4* +		2.88 [1.90,4.36]
All	6.07 [4.63-7.95]	1.34 [1.06,1.70]

OR = Odds Ratio; ORs compare GG to GT/TT

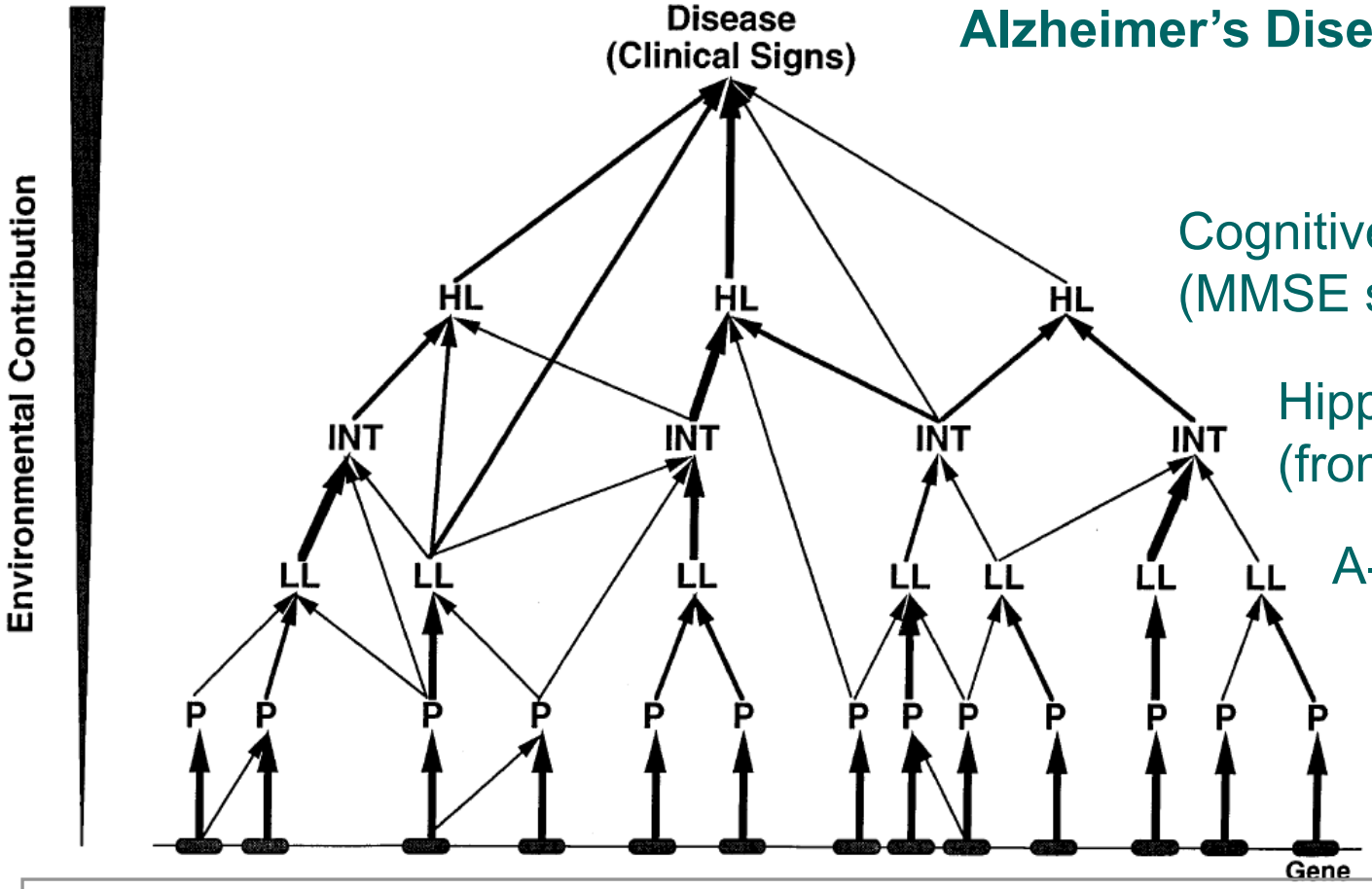
Suggests *GAB2* modifies LOAD risk in APOE e4 carriers

(Reiman et al, *Neuron* 2007; 54:713-720)

Intermediate Phenotypes

- Phenotype that is heritable, measurable, and has a closer relationship to the biological process involved in culmination of disease
- Represents a more elementary phenomenon
- The # of genes affecting intermediate phenotype variation is smaller than the number of genes affecting the full disease/trait phenotypic variation
- The genes affecting intermediate phenotypes have larger effect size
- **For an intermediate phenotype to be useful, it should be heritable & associated with disease/trait of interest!**

Intermediate Phenotypes



Alzheimer's Disease

Cognitive function (MMSE score)

Hippocampal atrophy (from MRI)

A-Beta protein level

P = Gene Product; LL = Low Level Phenotype; INT = Intermediate Phenotype; HL = High Level Phenotype

(Schork, NJ, Am J Respir Crit Care Med, 1997)

Lecture Outline

1) **Phenotype definition**
   1. Discrete verse quantitative traits
   2. Complex disease and natural history of disease
   3. Selecting your phenotype

2) **Measurement error**
   1. Properties of a good measure
   2. Consequences of measurement error

3) **Advantages of using standard phenotypes**
   1. Why is it important to use standard measures?
   2. Example of successful cross-study analyses
   3. Introduction to PhenX

2. Measurement Error

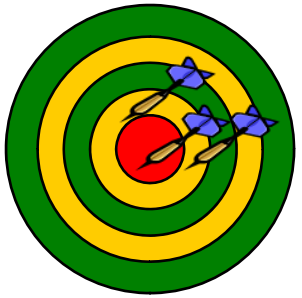
- Measure refers broadly to any way of capturing data on a certain characteristic of study subjects
- Method could be self-administered questionnaire, personal interview, physical exam, lab test, medical records extraction, etc…
- Regardless of characteristic or data collection method, there is a TRUE value of the characteristic for each study subject
- Any discrepancy between the TRUE value and the MEASURED value is **Measurement Error**

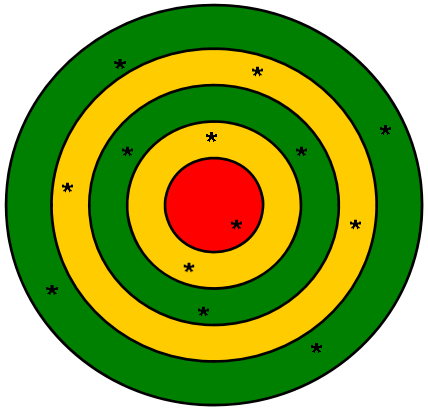
Properties of a good measure

- **Reliability**
  - describes consistency, reproducibility of a measurement
  - A good measurement should yield the same value if applied repeatedly under similar conditions
- **Validity**
  - describes accuracy of a measurement
  - A good measurement should yield the correct value/reflect the truth
- Reliability is a prerequisite for validity
- Reliability is necessary, but not a sufficient condition for validity

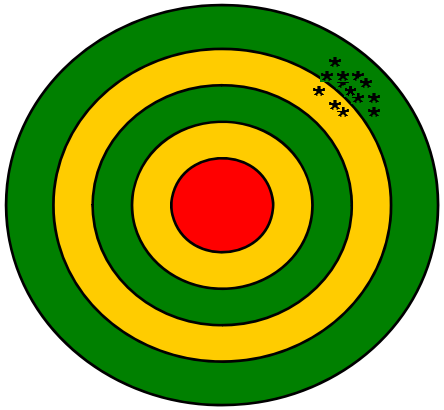
Properties of a good measure

- The goal is to hit the Bullseye with each dart:

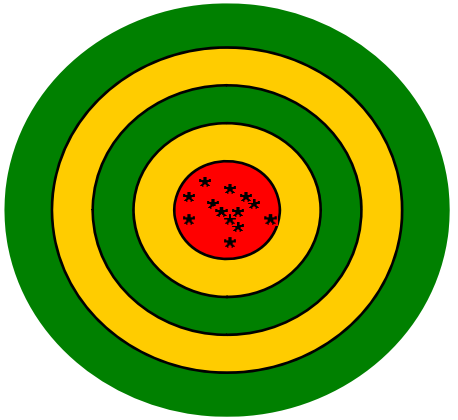




Results are neither reliable or valid



Results are reliable, but not valid



Results are both reliable and valid

Modified from D.S. Bhola, PhD

Quantifying Reliability

- Discrete/Categorical Traits
  - To what degree do the measurements agree beyond what we would expect by chance alone?
  - Kappa (κ) ranges from 0 to 1
  - Guidelines for Interpretation of Kappa (Source: Landis & Koch, 1977)
Kappa	Interpretation
> .80	Almost perfect
.61-.80	Substantial
.41-.60	Moderate
.21-.40	Fair
.00-.20	Slight

Kappa (κ)

Data layout for Calculating Kappa

Measure #1	Measure # 2 +	-	Total
+	a	b	a+b
-	c	d	c+d
Total	a+c	b+d	N

$$\kappa = P_o - P_e \,/\, 1-P_e$$

Where:

P_o = observed concordance (% agreement observed)

(a+d) / N

P_e = concordance expected by chance (% agreement expected by chance alone)

$$\left[\frac{(a+b)(a+c)}{N}\right] + \left[\frac{(b+d)(c+d)}{N}\right] \;/\; N$$

Kappa (κ): Example

Wright and colleagues (2000) studied genital-tract human papillomavirus (HPV) testing as possible screening test for cervical cancer. The examined agreement between test results on swabs obtained by clinicians with swabs obtained by screeners themselves. For 1415 women, both kinds of specimens were obtained:

Clinician collected	Self-collected +	Self-collected -	Total
+	170	132	302
-	128	985	1113
Total	298	1117	1415

P_o = observed concordance

(170+985) / 1415 = 0.816

P_e = concordance expected by chance

$$\frac{(302)(298)}{1415} + \frac{(1113)(1117)}{1415} = 0.666$$

κ = Po – Pe / 1- Pe

κ = 0.816 – 0.666 / 1-0.666

= 0.45, moderate agreement

(from Koepsell & Weiss, page 221)

Quantifying Validity

- True status of characteristic of interest must be known ("gold standard")
- Compare measure of your characteristic of interest to the gold standard

Data layout for assessing validity of binary test

Test Result	Condition present +	Condition present -
+	a	b
-	c	d
Total	a+c	b+d

a = # of true positives

b = # of false positives

c = # of false negatives

d = # of true negatives

Sensitivity = when condition truly present, how often does the test detect it?

= **a / (a+c)**

Specificity = when condition is truly absent, how often does test give a neg. result? = **d / (b+d)**

(from Koepsell & Weiss, page 223)

Consequences of Measurement Error

- Impact of measurement error on results depends on the way the error has arisen
- Measurement error of a discrete/binary outcome is termed misclassification
- Nondifferential (nonselective) misclassification of outcome
  - Present whether errors in assessing subject's status are similar regardless whether that subject has been exposed or not
  - Generally leads to an attenuation of the estimated size of a true association between exposure and disease
  - i.e. bias towards null
- Improving the resolution of measurement tools will allow more accurate characterization of the relationship between exposure (genotype) and disease!

Lecture Outline

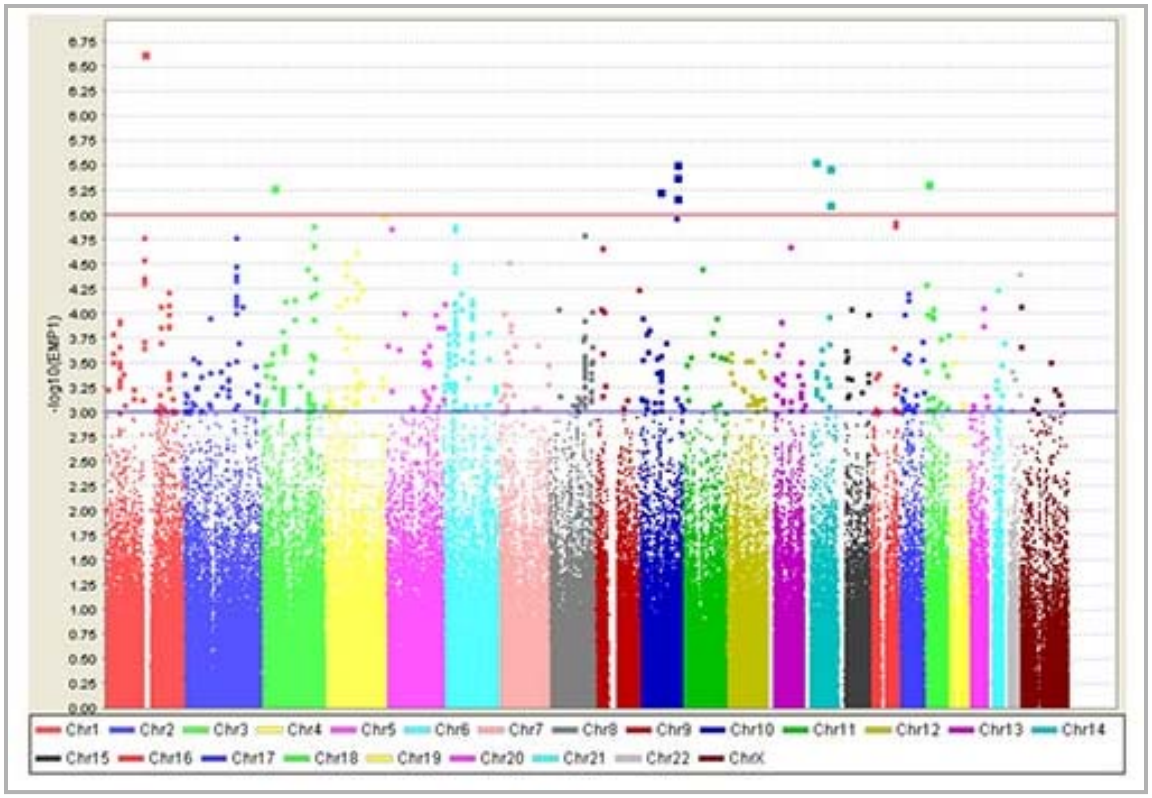
1) Phenotype definition
   1. Discrete verse quantitative traits
   2. Complex disease and natural history of disease
   3. Selecting your phenotype

2) Measurement error
   1. Properties of a good measure
   2. Consequences of measurement error

3) Advantages of using standard phenotypes
   1. Why is it important to use standard measures?
   2. Example of successful cross-study analyses
   3. Introduction to PhenX

3. Genome-Wide Association (GWA) Studies

- GWA studies measure > 100,000 single nucleotide polymorphisms (SNPs) across the genome & relate them to common diseases and traits
- Since 2005, over **160** GWA studies have identified robust SNP associations ($P < 10^{-7}$) for nearly **60** diseases and traits



- Type 2 Diabetes
- 386,731 markers

(http://www.broad.mit.edu/diabetes/scandinavs/type2.html)

Unique Aspects of GWA Studies

- Permit examination of genetic variation at an unprecedented level of resolution
- Allow “agnostic” genome-wide evaluation
- Once genome measured, can be related to any trait
- Most robust associations in GWAS reports have not been with genes previously suspected of being related to the disease
- Some significant associations are in regions that are not currently known to harbor genes

“The chief strength of the new approach also contains its chief problem: with more than 500,000 comparisons per study, the potential for false positive results is unprecedented.”

“Thus, the sine qua non for belief in any specific result from a GWAS is not the strength of the P value in the initial study, but the consistency and strength of the association across one or more large-scale replication studies.”

(Hunter DJ and Kraft P, NEJM, 2007)

Courtesy, Teri Manolio, NHGRI

Cross-Study Analysis is Essential

- More bang for the buck!
  - GWA and related studies are expensive
  - Combining studies increases ability to detect loci with moderate effect size
  - Once genome is characterized it can be related to traits beyond those focused on in the initial study (with appropriate consent)
- Potential for cross-study analysis limited by lack of standardized measures being included in GWAS
  - despite many risk factors common to multiple diseases (e.g. obesity, smoking, etc)

Association of rs1042725 (*HMGA2*) & height

Study (women only)	Mean Age	N	Mean height (cm) by genotype TT	CT	CC	*P*-value
GWA						
WTCCC (T2D)	57.9	792	160.4	161.5	162.2	0.0006
DGI (T2D)	65.2	638	160.0	161.3	162.1	0.003
DGI (Controls)	58.5	546	162.1	162.8	163.7	0.003
Combined		>4K				$4x10^{-8}$

(Weedon et al, Nature Genet 2007; 39:1245-50)

Association of rs1042725 (*HMGA2*) & height

Study (women only)	Mean Age	N	Mean height (cm) by genotype TT	CT	CC	*P*-value
GWA						
WTCCC (T2D)	57.9	792	160.4	161.5	162.2	0.0006
DGI (T2D)	65.2	638	160.0	161.3	162.1	0.003
DGI (Controls)	58.5	546	162.1	162.8	163.7	0.003
Combined		>4K				$4x10^{-8}$
Replication						
UKGCC T2D	64.0	820	159.0	159.3	159.9	0.037
EFSOCH parents	32.9	936	164.6	165.0	165.4	0.004
Combined		>19K				$3x10^{-11}$

(Weedon et al, Nature Genet 2007; 39:1245-50)

Association of rs1042725 (*HMGA2*) & height

Study (women only)	Mean Age	N	Mean height (cm) by genotype TT	CT	CC	*P*-value
GWA						
WTCCC (T2D)	57.9	792	160.4	161.5	162.2	0.0006
DGI (T2D)	65.2	638	160.0	161.3	162.1	0.003
DGI (Controls)	58.5	546	162.1	162.8	163.7	0.003
Combined		>4K				$4x10^{-8}$
Replication						
UKGCC T2D	64.0	820	159.0	159.3	159.9	0.037
EFSOCH parents	32.9	936	164.6	165.0	165.4	0.004
Combined		>19K				$3x10^{-11}$
All studies		>23K	effect size/C allele ~0.4cm			$4x10^{-16}$

(Weedon et al, Nature Genet 2007; 39:1245-50)

FTO Variant (rs9939609), T2 Diabetes, & Obesity

			Diabetes Association		
Study	Cases	Controls	OR	95% CI	*p*-value
WTCCC (TD2 Ph 1)	1,924	2,938	1.27	[1.16-1.37]	5×10^{-8}
WTCCC (TD2 Ph 2)	3,757	5,346	1.15	[1.09-1.23]	9×10^{-6}
- Adjusted for BMI			1.03	[0.96-1.10]	0.44

Association of rs9939609 with T2D risk mediated through BMI

			Mean BMI (kg/m^2)			
Study	% ♂	N	TT	AT	AA	*p*-value
WTCCC TD2 cases	58	1,913	30.2	30.5	32.0	8×10^{-6}
UKGCC TD2 cases	57	2,961	30.6	31.0	32.0	3×10^{-5}
EFSOCH controls	51	1,746	24.5	25.2	25.4	0.0002
EPIC-Norfolk (pop-based)	47	2,425	25.9	26.2	26.6	0.001
All studies (Bonferonni correction)						1.2x10-29

Frayling et al, Science 2007; 316:889-893

www.phenx.org

PhenX Domains

- Aging
- Alcohol, Tobacco, and Other Substances*
- Anthropometrics*
- Cancer
- Cardiovascular
- Central Nervous System
- Demographics*
- Child development
- Diet
- Diabetes
- Exposures & Responses
- Gastrointestinal
- Immunity
- Lung Function
- Ocular
- Oral Health
- Physical Activity
- Psychosocial
- Renal Function
- Reproduction
- Skin/Bone/Muscle

Demographic Measures

- Age
- Ancestry
- Race/Ethnicity
- Sex/Gender
- Current Marital Status
- Current Employment Status
- Education
- Income/Wealth
- Health Care
- Years in the U.S.

Summary Points

- Selecting appropriate phenotypes for your genomic research study is important
- Use reliable and valid measures to capture the information about your disease/trait and relevant covariates
- To increase potential for cross-study analysis, think about using commonly used measures with standard assessment protocols

Take home message…

PHENOTYPE,

PHENOTYPE,

PHENOTYPE!!!

Association of rs563694 & fasting glucose

Variations in the *G6PC2/ABCB11* genomic region are associated with fasting glucose levels

Wei-Min Chen,[1,2] Michael R. Erdos,[3] Anne U. Jackson,[4] Richa Saxena,[5] Serena Sanna,[4,6] Kristi D. Silver,[7] Nicholas J. Timpson,[8] Torben Hansen,[9] Marco Orrù,[6] Maria Grazia Piras,[6] Lori L. Bonnycastle,[3] Cristen J. Willer,[4] Valeriya Lyssenko,[10] Haiqing Shen,[7] Johanna Kuusisto,[11] Shah Ebrahim,[12] Natascia Sestu,[13] William L. Duren,[4] Maria Cristina Spada,[6] Heather M. Stringham,[4] Laura J. Scott,[4] Nazario Olla,[6] Amy J. Swift,[3] Samer Najjar,[13] Braxton D. Mitchell,[7] Debbie A. Lawlor,[8] George Davey Smith,[8] Yoav Ben-Shlomo,[14] Gitte Andersen,[9] Knut Borch-Johnsen,[9,15,16] Torben Jørgensen,[15] Jouko Saramies,[17] Timo T. Valle,[18] Thomas A. Buchanan,[19,20] Alan R. Shuldiner,[7] Edward Lakatta,[13] Richard N. Bergman,[20] Manuela Uda,[6] Jaakko Tuomilehto,[18,21] Oluf Pedersen,[9,16] Antonio Cao,[6] Leif Groop,[10] Karen L. Mohlke,[22] Markku Laakso,[11] David Schlessinger,[13] Francis S. Collins,[3] David Altshuler,[5] Gonçalo R. Abecasis,[4] Michael Boehnke,[4] Angelo Scuteri,[23,24] and Richard M. Watanabe[20,25]

- Rationale: Understanding genetic variants that regulate fasting glucose concentrations may further our understanding of the pathogenesis of diabetes

(Chen, W *et al*., JCI, July, 2008)

Association of rs563694 & fasting glucose*

Study	n	Mean fasting glucose (mM) CC	AC	AA	Effect Size (mM)	*p*-value
GWA						
FUSION stage I	1,233	5.26	5.31	5.33	0.051	8.0×10^{-4}
SardiNIA	3,855	4.88	4.95	5.00	0.064	7.6×10^{-5}
GWA meta analysis						3.5×10^{-7}

* In non-diabetic individuals

(Chen, W *et al*., JCI, July, 2008)

Association of rs563694 & fasting glucose*

Study	n	Mean fasting glucose (mM) CC	AC	AA	Effect Size (mM)	*p*-value
GWA						
FUSION stage I	1,233	5.26	5.31	5.33	0.051	8.0×10^{-4}
SardiNIA	3,855	4.88	4.95	5.00	0.064	7.6×10^{-5}
GWA meta analysis						3.5×10^{-7}
Follow-up						
FUSION stage II	655	5.28	5.44	5.46	0.068	2.0×10^{-3}
Amish	1,655	4.90	4.89	5.03	0.090	4.1×10^{-5}
METSIM	4,386	5.55	5.64	5.71	0.145	1.3×10^{-10}
Follow-up meta analysis						6.3×10^{-28}
Overall meta analysis						6.1×10^{-35}

Concluded that G6PC2, a glucose-6-phosphatase (expressed in pancreatic cells), may underlie variation in fasting glucose