Nutritional Epidemiology in the Genomic Age

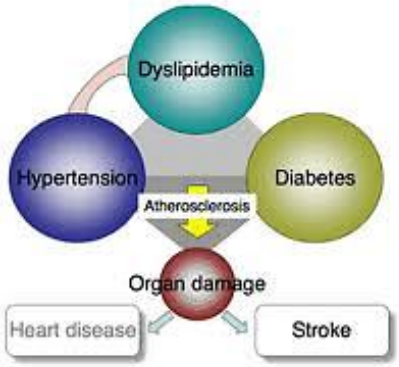
Saroja Voruganti, PhD
Assistant Professor
Department of Nutrition and
UNC Nutrition Research Institute

UNC NUTRITION RESEARCH INSTITUTE

Learning objectives

- Types of genetic association approaches and their relevance to nutritional research

- Models used to analyze associations of genetic variants with disease phenotypes and gene-nutrient interactions



Gene X Gene

Gene X Environment (diet)

Environmental (diet) Factors

Genetic Factors

UNC NUTRITION RESEARCH INSTITUTE

Research question??

- What are the genes that affect nutrient metabolism?

Or

- How do our nutrient or diet intake affect the expression of a gene?

Nutritional Epidemiological approaches

- Correlation studies
- Special exposure groups
- Migrant studies
- Case control and cohort studies
- Controlled trials

Willett W. Overview of Nutritional Epidemiology
gDOI:10.1093/acprof:oso/9780199754038.003.0001

Genetic Epidemiological approaches

- Case studies
- Cross sectional studies
- Cohort
- Case-control
- Family-based, twin and trio studies
- Clinical trials

Defining the phenotype and genotype

- Phenotype
  - Is it properly defined?
  - Is it genetically controlled?
  - Is it likely to have effects mediated by a given environmental factor?

- Genotype
  - Does it show evidence for linkage, association, or interaction?
  - Are the SNPs in promoter, intron or exon?
  - What do we know about the SNP?

Biological samples and other data

- Nutrient data
- DNA (from blood, saliva and other tissues)- genotyping or sequencing, epigenetics
- RNA (from blood or tissues) – transcriptomic profiles
- Blood (serum or plasma) –biochemical variables, metabolomics, proteomics, lipidomics, etc
- Urine – biochemical variables, metabolomics
- Data such as age, sex, BMI, etc
- Medical history
- Environmental factors

How can variants affect phenotypes?

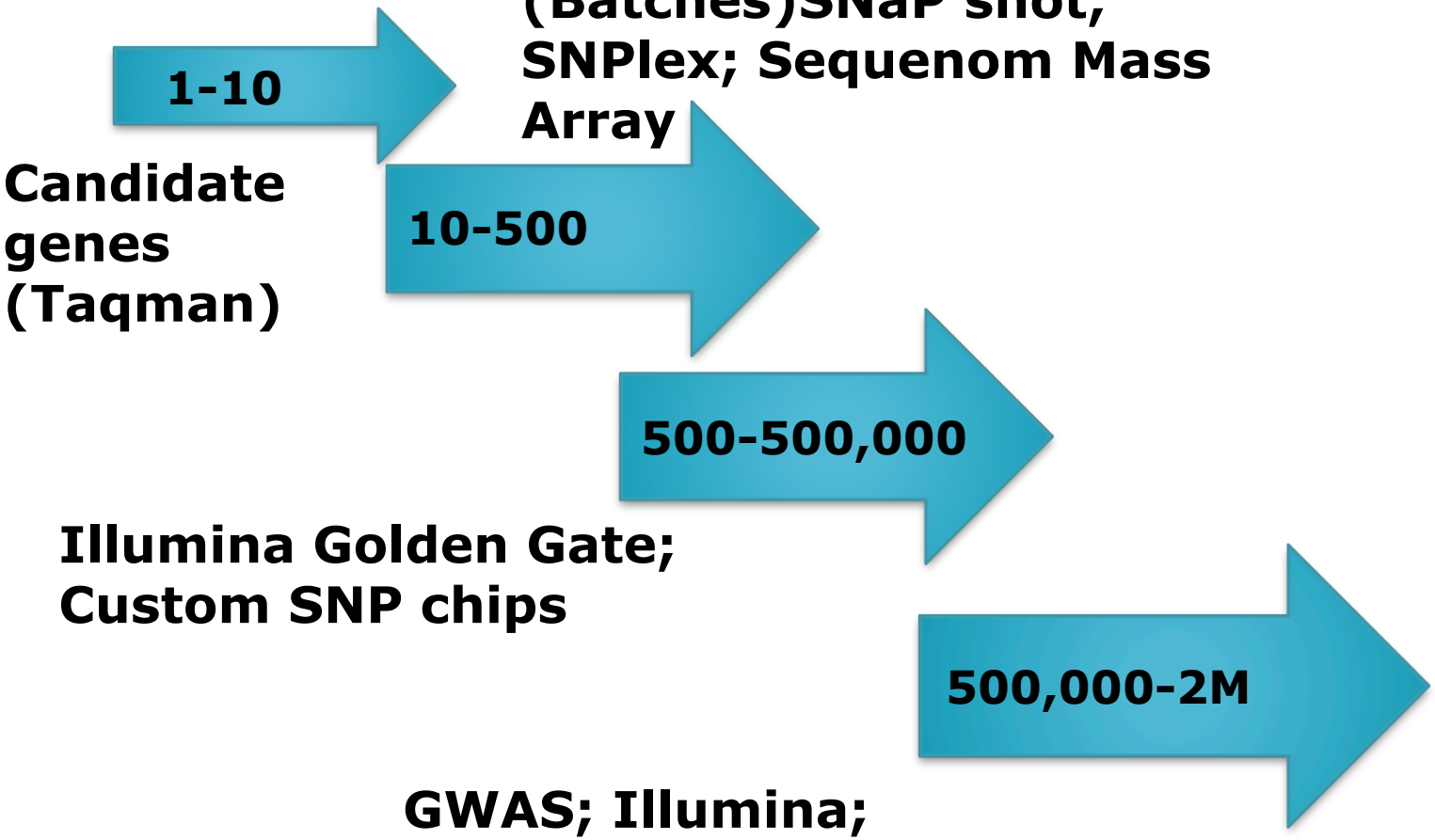
- silent; most variants have no effect
- Altered protein sequence – nonsynonymous, nonsense, splice, stop loss
- Altered RNA processing
- Altered RNA expression (regulatory)
- Other

Approaches to genotyping

- Candidate genes:
- genotype only markers in genes potentially related to the trait

- Genome screen:
- genotype anonymous markers spanning the genome at regular intervals

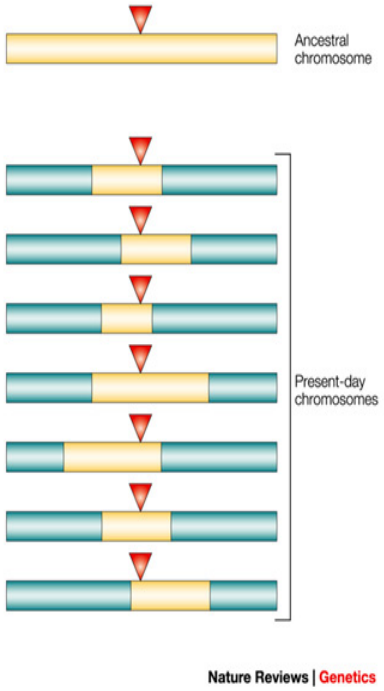
Genotyping



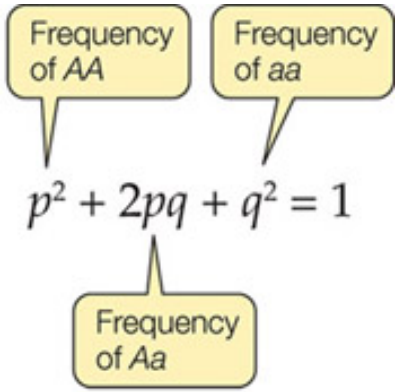
Adapted from Edenberg and Liu. Cold Spring Harbor Laboratory Press; 2009

Terminology

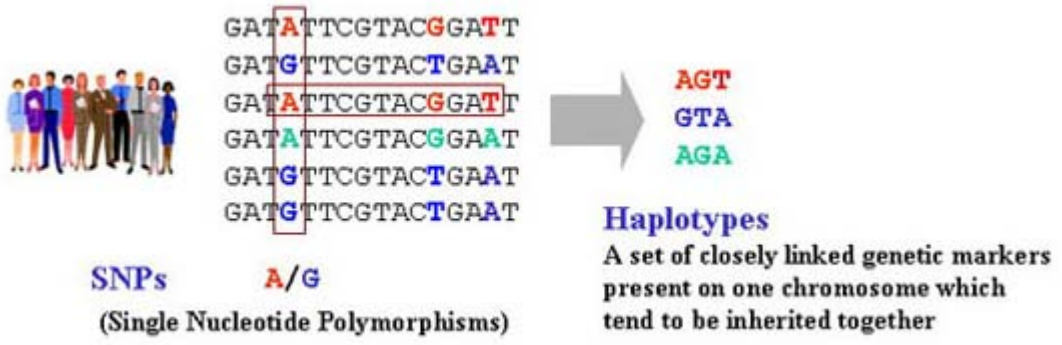
Linkage Disequilibrium



Hardy-Weinberg Principle



Haplotype



Ardlie, Kruglyak & Seielstad
Nature Reviews Genetics, 2002; 3, 299-309

Genome-wide Approaches

LINKAGE

Design:Family
Phenotypes:Quantitative /Qualitative Phenotypes
Markers:STR or SNP
Information:Segregation (IBD)

ASSOCIATION

Design:Case-Control / Family
Phenotypes:Quantitative /Qualitative Phenotypes
Markers:STR or SNP
Information:Linkage Disequilibrium (IBS)

Modeling in genetic epidemiology

- Homozygote (AA) –
- 2 copies of major allele ('common')

- Heterozygote (Aa) –
- 1 copy of major allele and 1 of minor allele

- Homozygote (aa) –
- 2 copies of minor allele ('variant')

Modeling in genetic epidemiology

- the mode of inheritance
  - 1. Additive
  - 2. Dominant
  - 3. Recessive

- With family data/ pedigrees – assess mode of inheritance

- BUT….Can't be done in studies of unrelated individuals, complex with common traits
- Statistical power is reduced if you specify the wrong model

Recoding for alternative models

	Additive	Dominant	Recessive
AA	0	0	0
AG	1	1	0
GG	2	1	1

- Dominant model combines AG+GG
  - Only need one copy of rare allele for disease
  - Used when frequency of GG is low
- Recessive model combines AA & AG
  - Have to have 2 copies of rare allele (G) for disease
  - Rarely used

Power calculation

- http://bmcgenet.biomedcentral.com/articles/10.1186/1471-2156-9-36 (PGA-Power calculator for case-control genetic association studies)
- http://www.biostat.ucsf.edu/sampsize.html
- http://homepage.stat.uiowa.edu/~rlenth/Power/
- http://biomath.info/power/
- http://pngu.mgh.harvard.edu/~purcell/gpc/

Genetic association tools

- http://goldenhelix.com/products/SNP_Variation/index.html
- http://genemapping.org/online material/online-resources
- http://www.broadinstitute.org/scientific-community/software?criteria=Genetic%20Analysis
- http://bmcresnotes.biomedcentral.com/articles/10.1186/1756-0500-4-158
- http://www.biostat.wustl.edu/genetics/geneticssoft/SoftwareList.htm

Genetic association tools

- http://www.stats.ox.ac.uk/~marchini/software/gwas/gwas.html
- http://www.disgenet.org/web/DisGeNET/menu;jsessionid=16q535dpjpour10rfwtudqdjt4
- http://biostats.usc.edu/software
- http://pngu.mgh.harvard.edu/~purcell/

XWAS: A Software Toolset for Genetic Data Analysis and Association Studies of the X Chromosome

Feng Gao*, Diana Chang*, Arjun Biddanda*, Li Ma, Yingjie Guo, Zilu Zhou, and Alon Keinan

Websites

- UCSC Genome Browser – https://genome.ucsc.edu/cgi-bin/hgGateway

- NCBI Map Viewer http://www.ncbi.nlm.nih.gov/projects/mapview/

Others

- Online Mendelian Inheritance in Man
  http://www.ncbi.nlm.nih.gov/omim

- Gene
  http://www.ncbi.nlm.nih.gov/gene

- Gene Cards
  http://www.genecards.org/

Others

- HAPMAP

http://hapmap.ncbi.nlm.nih.gov/

- ENCODE

https://genome.ucsc.edu/ENCODE/

Or

http://www.genome.gov/10005107

LocusZoom

- http://locuszoom.sph.umich.edu/locuszoom/
- a tool to plot regional association results from genome-wide association scans or candidate gene studies. This is Version 1.1

Effects of gene-environment interaction on phenotypes

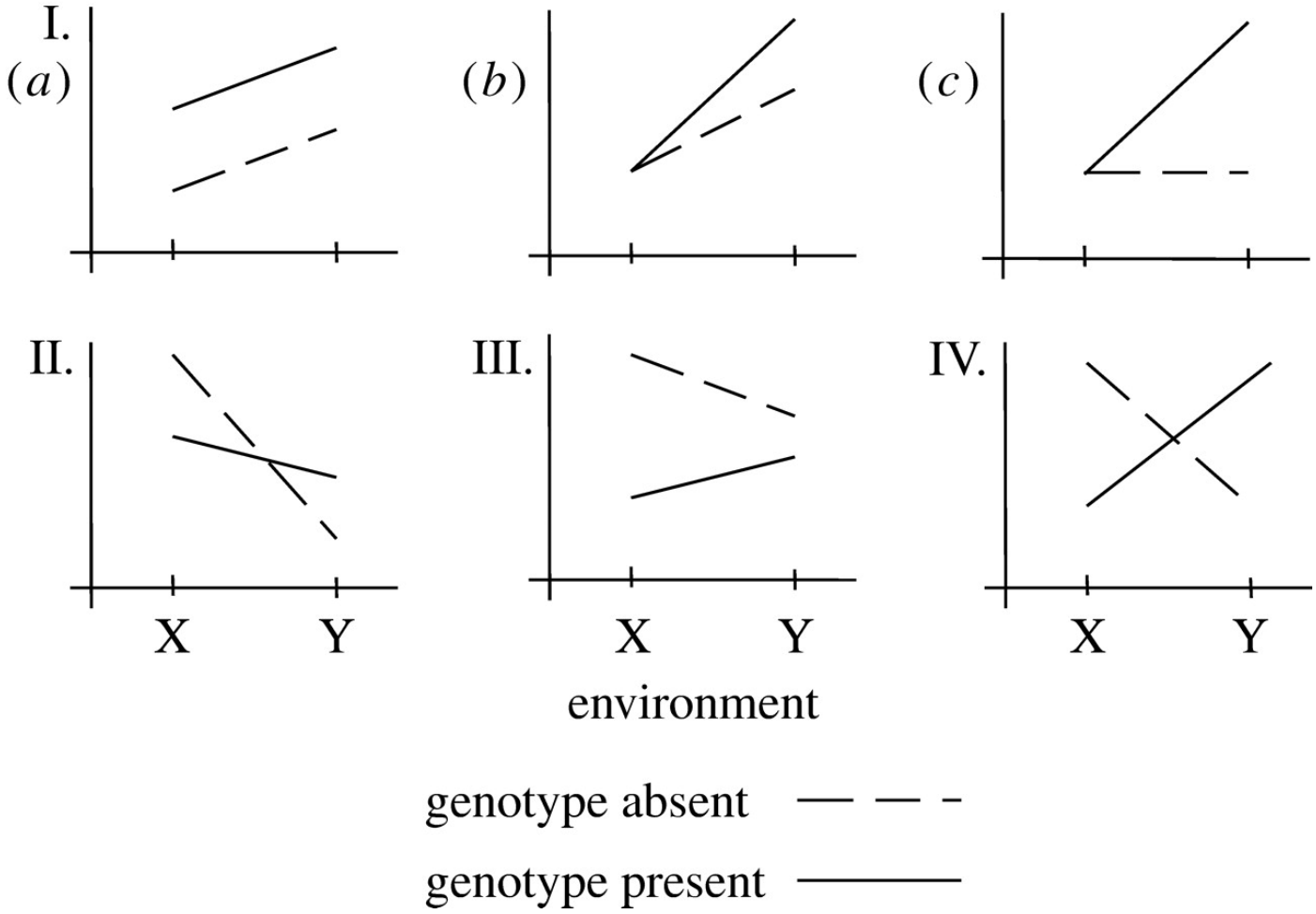
What is Gene-Environment Interaction (GEI)?

- Distinct effects of an environmental factor in individuals with different genotypes

or

- Distinct effects of a genotype in two different environments

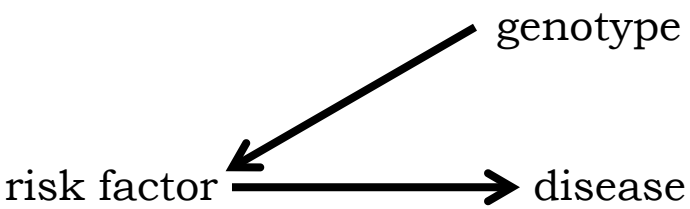
Gene-Environment Interaction



Kraft and Hunter.. 2005. Philosophical Transactions of the Royal Society B.

Gene by nutrient interaction effects on metabolic disease

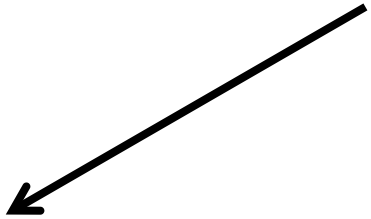
Model I -Phenylketonuria



Model 1

Genotype increases the expression of risk factor

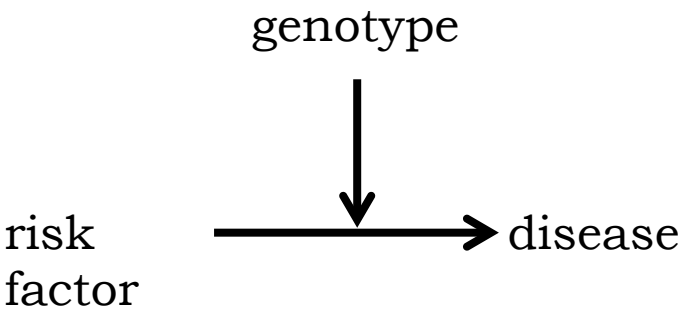
Mutation in phenylalanine hydroxylase



High levels of phenylalanine in blood

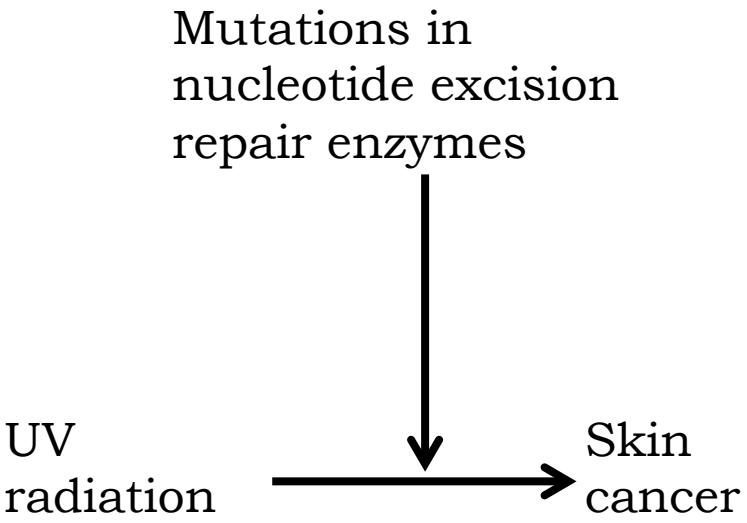
PKU

Model II – Xeroderma Pigmentosum

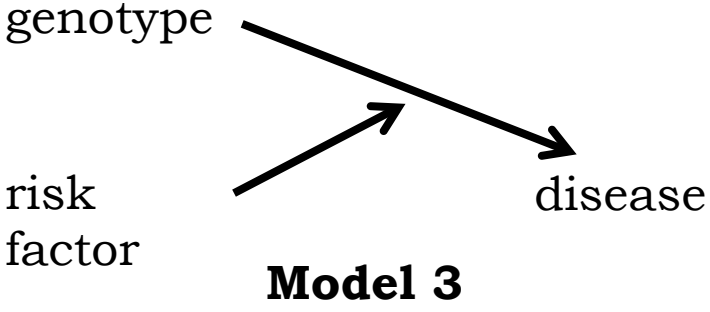


Model 2

Genotype excarbates the effect of the risk factor

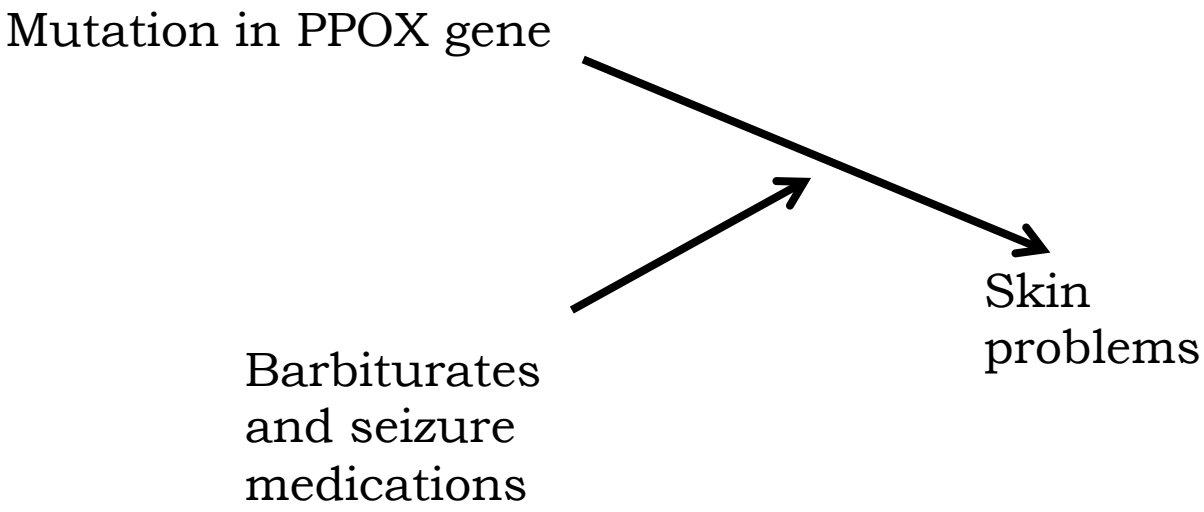


Model III- Porphyria variegate

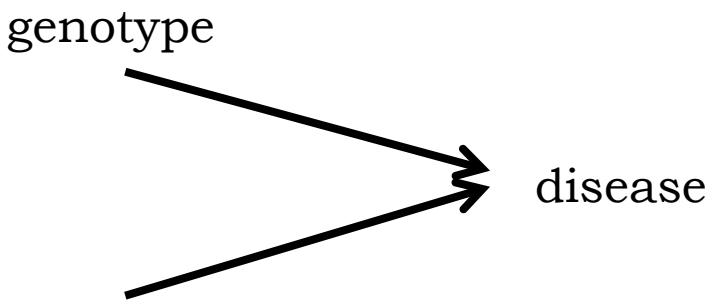


Model 3

The risk factor excarbates the effect of the genotype

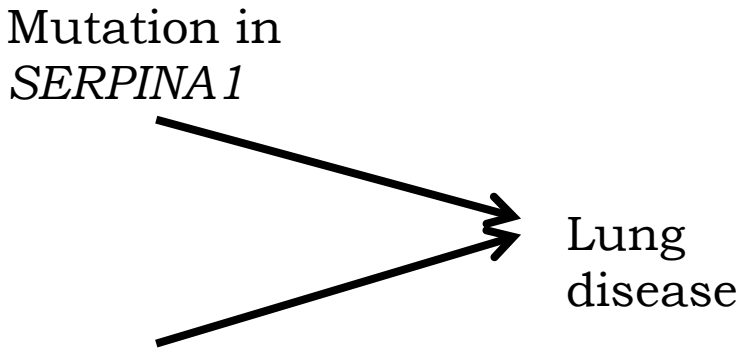


Model IV- alpha-1 antitrypsin deficiency



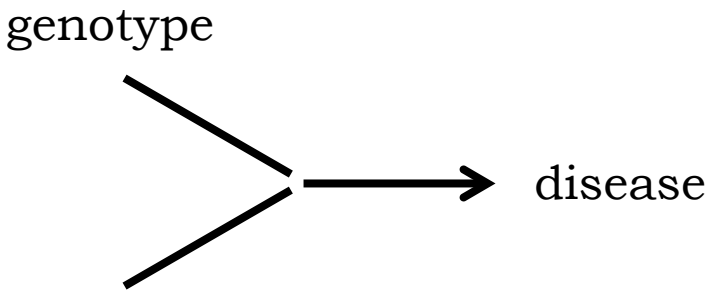
Genotype and risk factor each influence the risk by themselves

Model 4



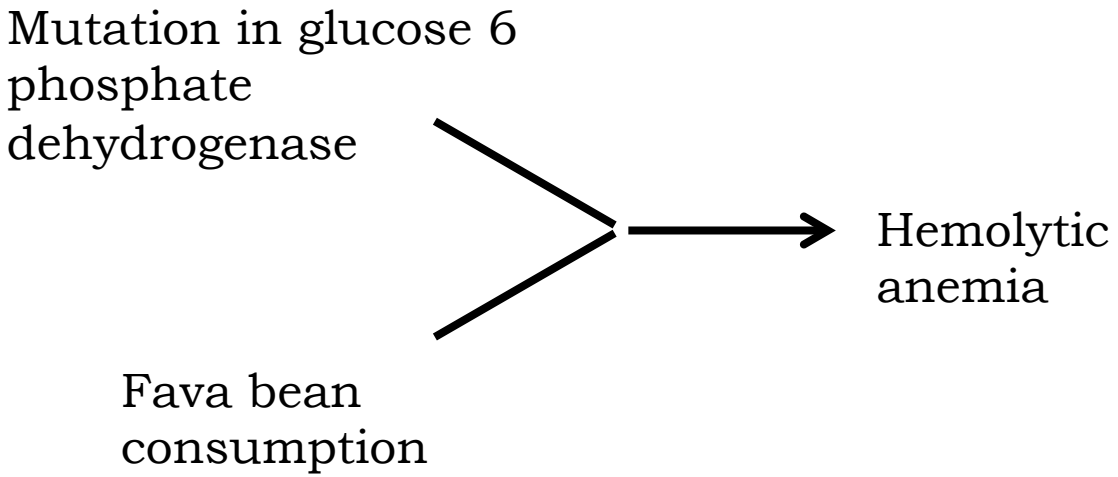
Model 4

Model V-G6PD deficiency



Both genotype and risk factor are required to raise the risk

Model 5



Nutrigenetic differences

- Most of them may have been inherited from our ancestors
- Genetic variation affects food tolerances among populations
- Nutritional environments seem to be the major determinants of human variation evolution
- Populations vary in their requirements for foods and response to diet

SNP by Environment Interaction

Main effects model:

- $T(E) = \beta^{M}_{0i} + \beta^{M}_{1i}E + \beta^{M}_{2i}SNP$

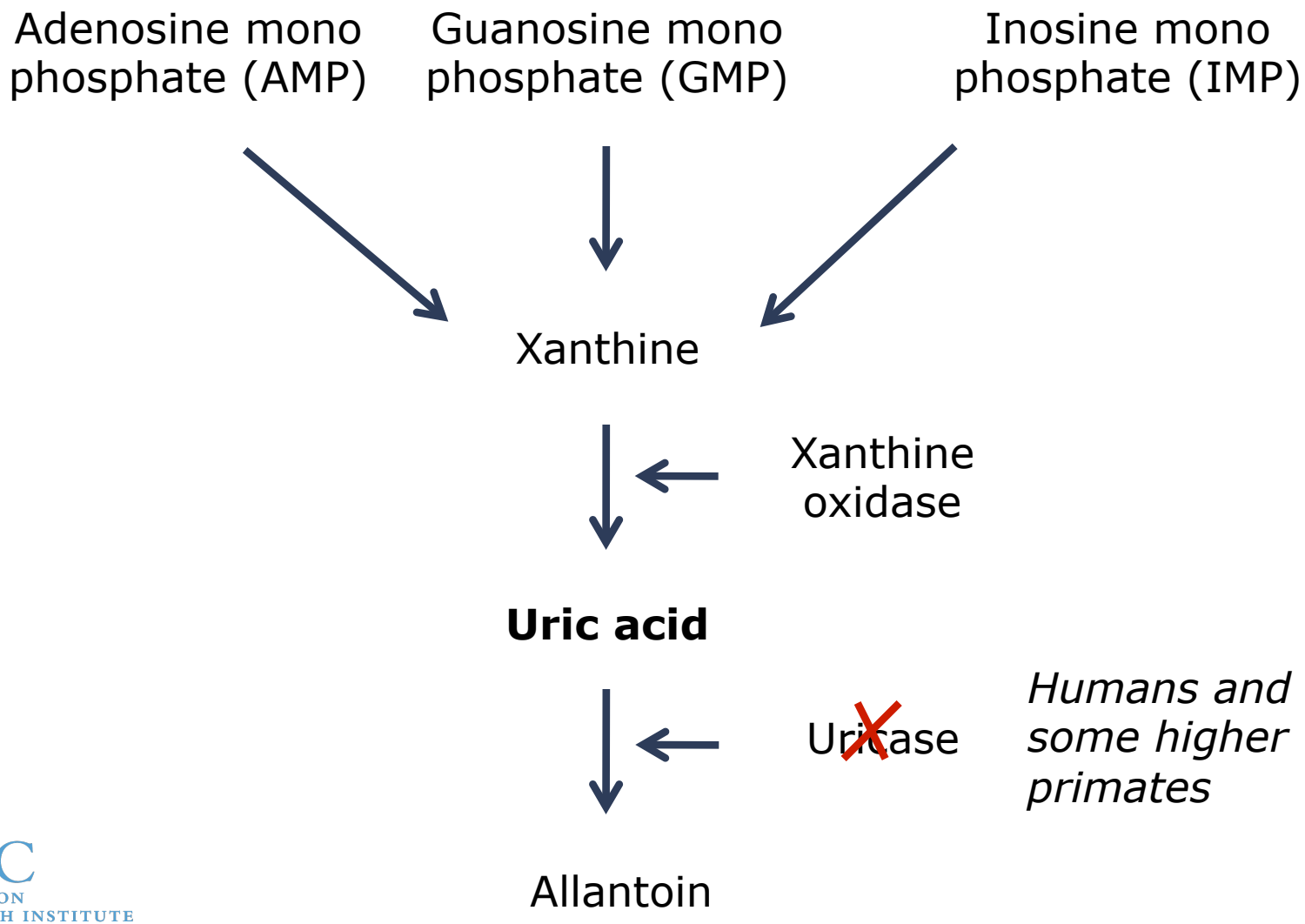
Interaction effects model:

- $T(E) = \beta^{I}_{0i} + \beta^{I}_{1i}E + \beta^{I}_{2i}SNP + \beta^{I}_{3i}SNPxE$
  - $T(E)$ = variation in the phenotype T,
  - β^{M} = coefficients related to main effects,
- β^{M} = coefficients related to interaction effects,
  - E = environmental factor,
- SNP is usually coded as 0,1 and 2 based on the number of rare alleles, and
  - SNP x E= interaction term

Example

UNC
NUTRITION
RESEARCH INSTITUTE

Serum uric acid



Strong Heart Family study (SHFS) [PI: Dr. Shelley Cole]

- Is a genetic study of CVD risk in American Indians

- It is the genetic component of the Strong Heart Study started in 1998

- More than 3800 members from multigenerational families enrolled from three centers located in Arizona, Dakotas and Oklahoma

UNC NUTRITION RESEARCH INSTITUTE

Viva La Familia [PI: Dr. Nancy Butte]

- Overweight/obese Hispanic children aged 4-19 years were recruited

- Some unique phenotypes such as calorimetry measurements, physical activity and energy expenditure have been collected

- Genome-wide SNP, exome and metabolomic data available

UNC NUTRITION RESEARCH INSTITUTE

Descriptives

	SHFS	VFS
Age	39.50 ± 17	11.0 ± 4
Serum uric acid (mg/dl)	5.1 ± 1.5	5.2 ± 1.7
Hyperuricemia (%)	17	25
Sugars intake (% of total calories)	16.3	22
Heritability (%)	46	45

SLC2A9* SNPs and serum uric acid levels (SHFS)**

	All			Arizona			Dakotas			Oklahoma		
SNP	Risk allele/ frequency	Effect size (%)[a]	P-value	Risk allele/ frequency	Effect size (%)	P-value	Risk allele/ frequency	Effect size (%)	P-value	Risk allele/ frequency	Effect size (%)	P-value
rs16890979	A/0.47	4.6	1.3E−31	A/0.55	3.8	3.8E−11	A/0.42	3.2	1.1E−8	A/0.43	5.1	7.8E−13
Post. Prob[b]		0		0.12			0			0		
rs6832439	A/0.47	4.5	7.7E−31	A/0.55	3.8	2.8E−11	A/0.41	3.3	1.5E−8	A/0.47	4.8	7.2E−12
Post. Prob[b]		0		0.09			0			0		
rs6449213	G/0.28	4.5	1.5E−29	G/0.37	3.0	1.6E−11	G/0.20	3.3	2.8E−9	G/0.27	6.4	1.2E−14
Post. Prob[b]		1.0		0.10			1.0			1.0		
rs13131257	A/0.48	4.5	1.8E−29	A/0.56	4.2	5.8E−12	A/0.42	3.3	5.0E−8	A/0.46	4.1	2.6E−10
Post. Prob[b]		0		0.43			0			0		
rs737267	A/0.49	4.4	2.9E−29	A/0.56	4.1	9.1E−12	A/0.45	3.1	4.5E−8	A/0.48	4.4	1.1E−10
Post. Prob[b]		0		0.21			0			0		
rs10805346	A/0.29	4.1	5.4E−28	A/0.21	4.1	2.4E−11	A/0.39	3.3	1.4E−8	A/0.34	2.7	1.6E−8
Post. Prob[b]		1.0		1.0			1.0			0		
rs12498956	C/0.47	3.4	5.1E−23	C/0.43	3.3	5.9E−10	C/0.49	2.1	1.0E−5	C/0.49	4.1	1.1E−9
Post. Prob[b]		0		0.05			0			0		

[a]Effect size – proportion of residual phenotypic variance that is explained by the SNP.
[b]Post. Prob – posterior probability of a functional effect. Posterior probability of >0.95 is considered as statistically significant evidence of functional effect.
Final model included covariates age, sex, age × sex, body mass index (BMI), eGFR, type 2 diabetes status, self-reported alcohol intake and medications.

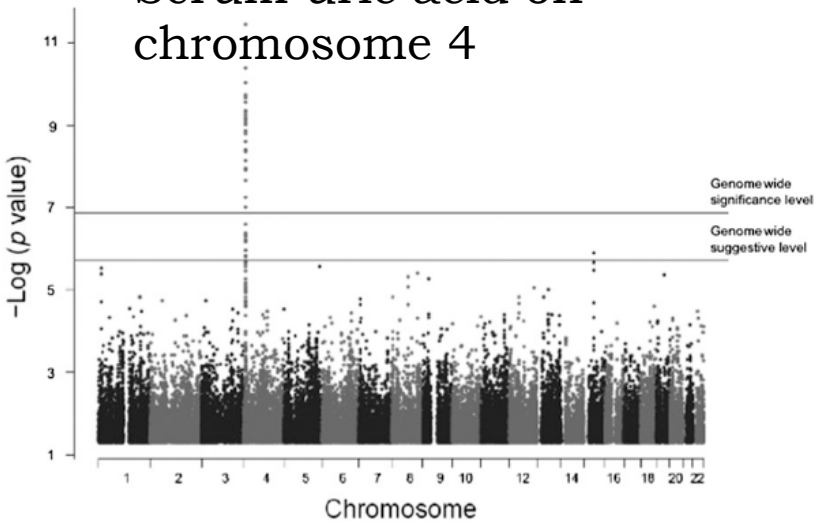
Dietary variable	All		Arizona		Dakotas		Oklahoma	
	β (SE)	P value	β (SE)	P value	β (SE)	P value	β (SE)	P value
Alcohol intake	-0.219 (0.03)	5.6×10^{-10}	-0.187 (0.04)	4.1×10^{-6}	-0.175 (0.04)	9.0×10^{-5}	-0.192 (0.04)	3.2×10^{-6}
Protein intake	0.0007 (0.0002)	0.0004	0.0004 (0.0002)	0.16	0.0008 (0.0003)	1.8×10^{-2}	0.0012 (0.0004)	0.008
Simple sugars	0.0003 (0.0002)	0.82	-0.0009 (0.0003)	0.72	0.0014 (0.0003)	0.65	0.0007 (0.0004)	0. 768

UNC NUTRITION RESEARCH INSTITUTE

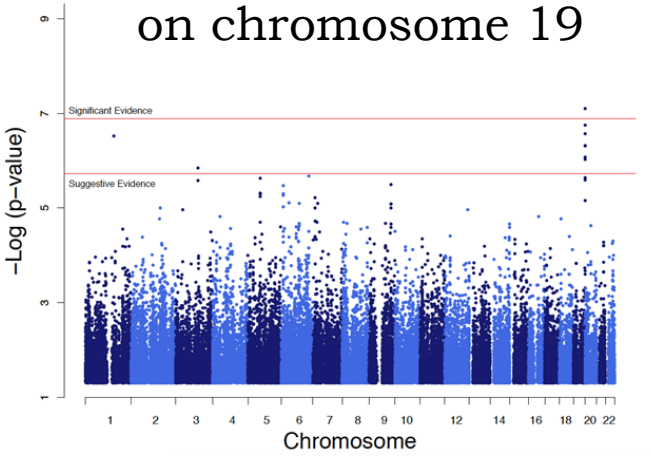
* Solute carrier family 2, member 9
** Voruganti et al., EJHG 2014

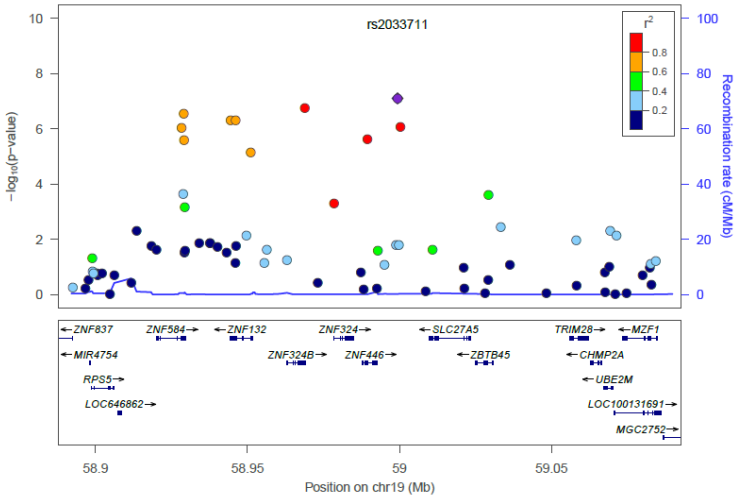
Genetic influence on serum uric acid and clearance

Serum uric acid on chromosome 4



Uric acid clearance on chromosome 19

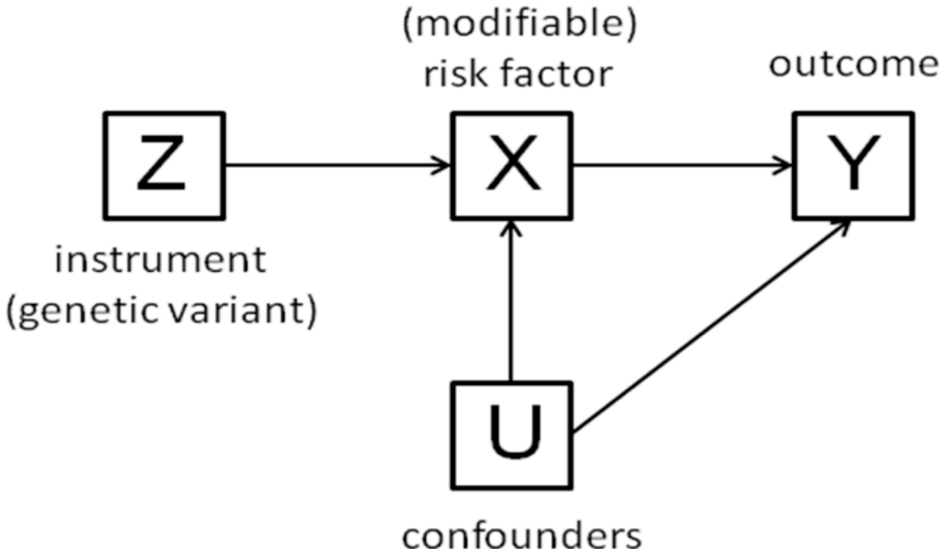




Locus Zoom plot showing the most significant SNPs on chr 19q13

UNC NUTRITION RESEARCH INSTITUTE

Mendelian randomization



SLC2A9 variants
(Instrument)

Serum uric acid
(risk factor)

Age, sex, age*sex
(Confounders)

Chronic kidney disease
(Outcome)

Association of SUA genetic risk score with kidney function markers

Study	Phenotype	B coefficient	SE	P value
Strong Heart Family study - All	Serum uric acid	0.0561	0.004	**3.5×10^{-38}**
	GFR	-0.0099	0.005	**0.04**
	UACR	-0.0095	0.005	0.05
	Serum creatinine	0.0202	0.005	**5.0×10^{-5}**
	GFR	-0.0165	0.009	0.05
	UACR	0.0085	0.008	0.24
	Serum creatinine	0.0152	0.008	0.07
Dakotas	Serum uric acid	0.0609	0.005	**1.0×10^{-33}**
	GFR	-0.0152	0.005	**3.0×10^{-3}**
	UACR	-0.0096	0.005	0.05
	Serum creatinine	0.0205	0.005	**4.0×10^{-5}**
Oklahoma	Serum uric acid	0.0602	0.005	**1.8×10^{-33}**
	GFR	-0.0144	0.005	**0.004**
	UACR	-0.009	0.005	0.06
	Serum creatinine	0.0198	0.005	**7.0×10^{-5}**
San Antonio Family Heart Study	Serum uric acid	0.0799	0.014	**4.7×10^{-8}**
	GFR	0.0018	0.013	0.89
	UACR	0.0456	0.013	**6.0×10^{-4}**
	Serum creatinine	-0.0106	0.013	0.42
Genetics of Kidney Disease in Zuni Indians	Serum uric acid	0.0587	0.010	**9.8×10^{-9}**
	GFR	-0.017	0.010	0.10
	UACR	0.0064	0.010	0.53
	Serum creatinine	0.0185	0.010	0.07
Genetics of Coronary Artery Disease in Alaska Natives	Serum uric acid	0.0361	0.009	**6.0×10^{-5}**
	GFR	-0.0116	0.013	0.37
	UACR	-0.0204	0.013	0.12
	Serum creatinine	0.0114	0.011	0.30

Dietary Factors affecting serum uric acid levels

Fructose
[Carbonated beverages, most canned products, honey]

High-purine foods and amino acids
[Organ meats such as liver, spleen, heart etc]

Alcohol

ATP depletion

AMP, GMP or IMP

Dehydration

Competition with uric acid for the same transporter (SLC2A9)

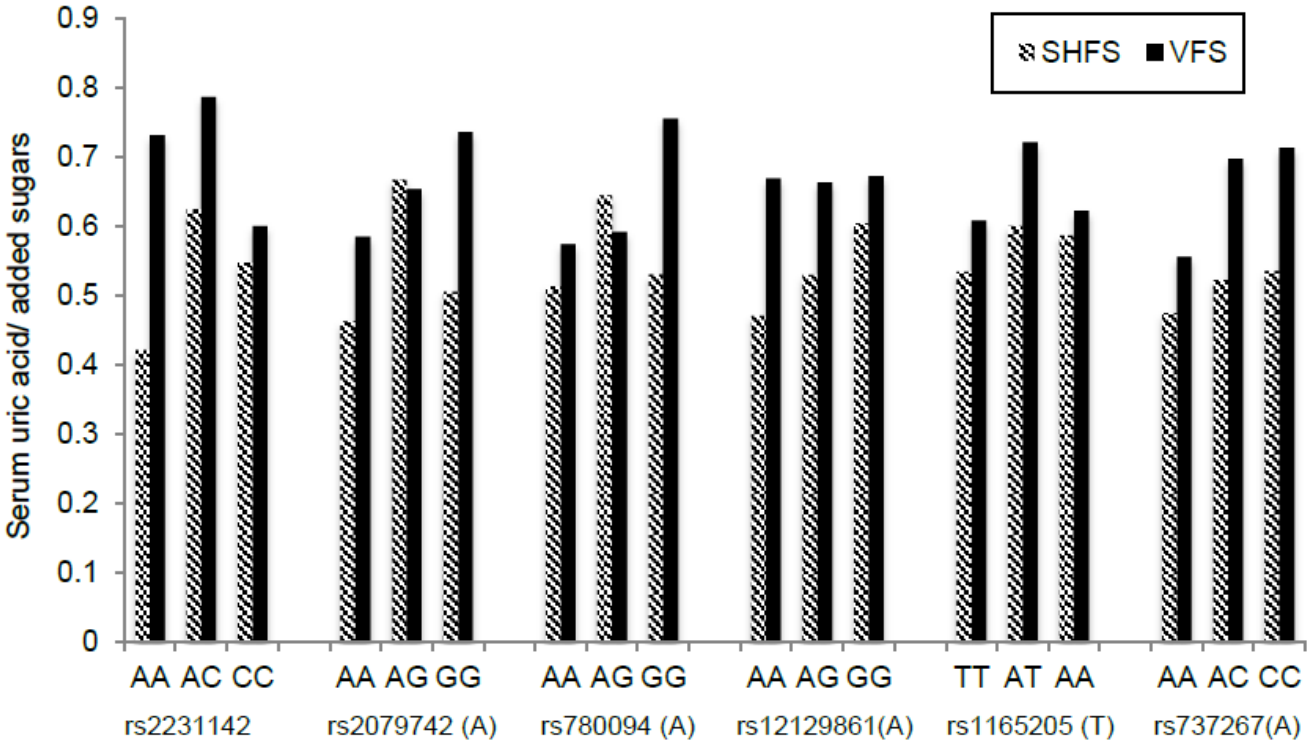
Hyperuricemia

UNC NUTRITION RESEARCH INSTITUTE

Uric acid and Fructose

- Uric acid is a byproduct of fructose degradation and shares a transporter with fructose (GLUT9/SLC2A9)
- Fructokinase is poorly regulated and phosphorylates fructose rapidly
- Fructose upregulates its transporter GLUT5 as well as fructokinase
- Serum uric acid increases rapidly after ingestion of fructose
- Fructose interferes with uric acid excretion

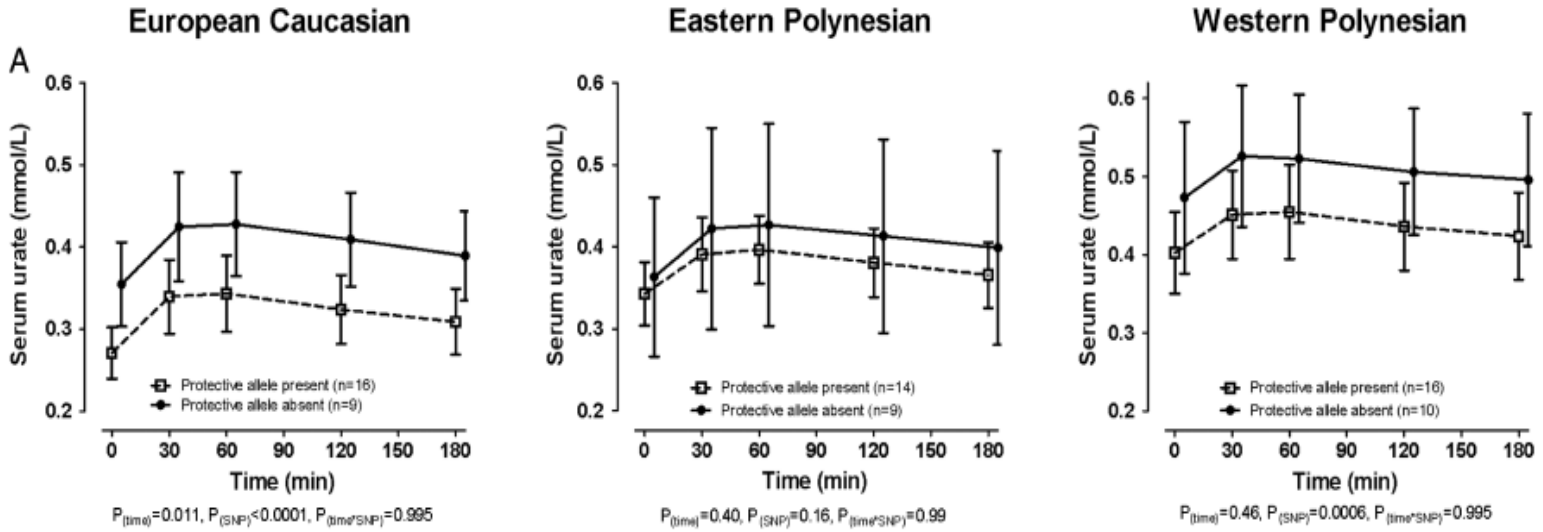
Genotype-specific differences in SUA/added sugars



Minor allele shown next to the SNP in parantheses; added sugars are shown as percent of calories

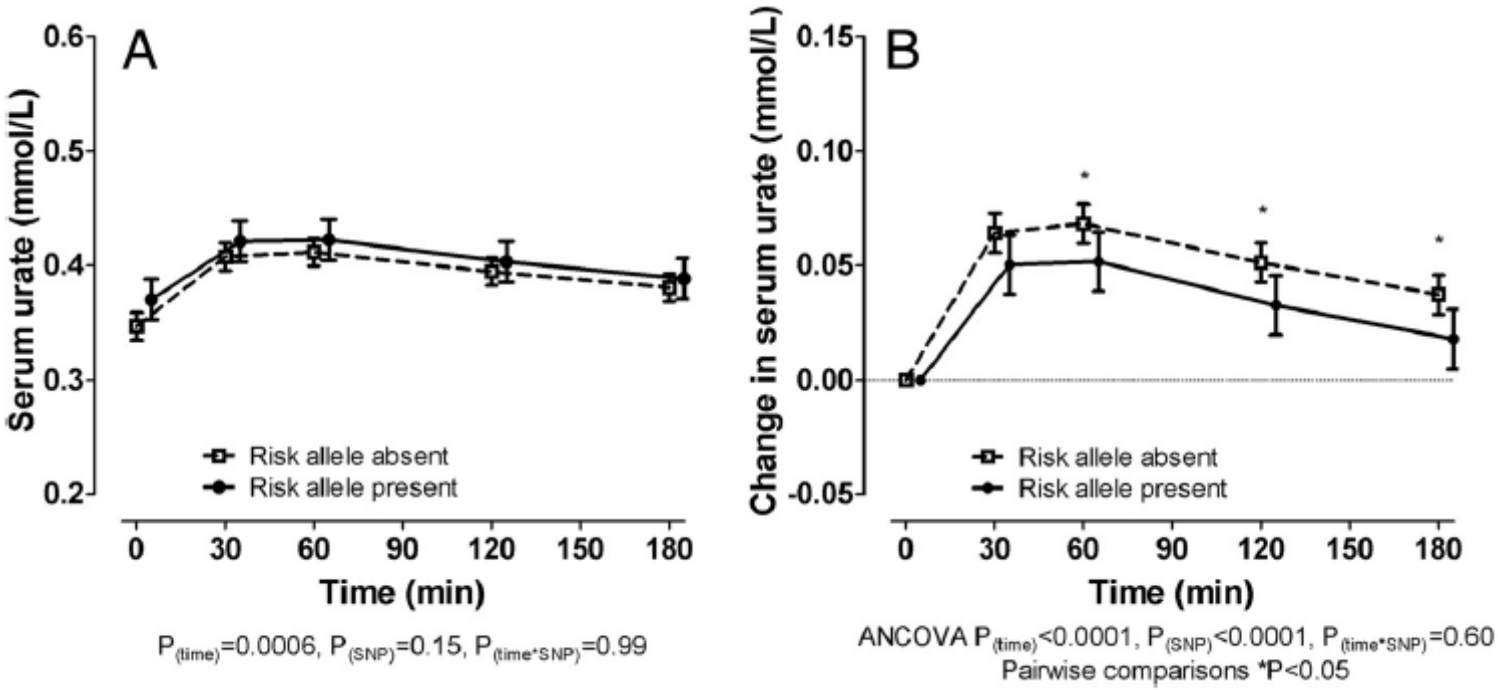
UNC NUTRITION RESEARCH INSTITUTE

Population-specific effects of *SLC17A1* on serum uric acid concentrations during a fructose load



Dalbeth et al. Ann Rheum Dis. 2014; 73: 313-314

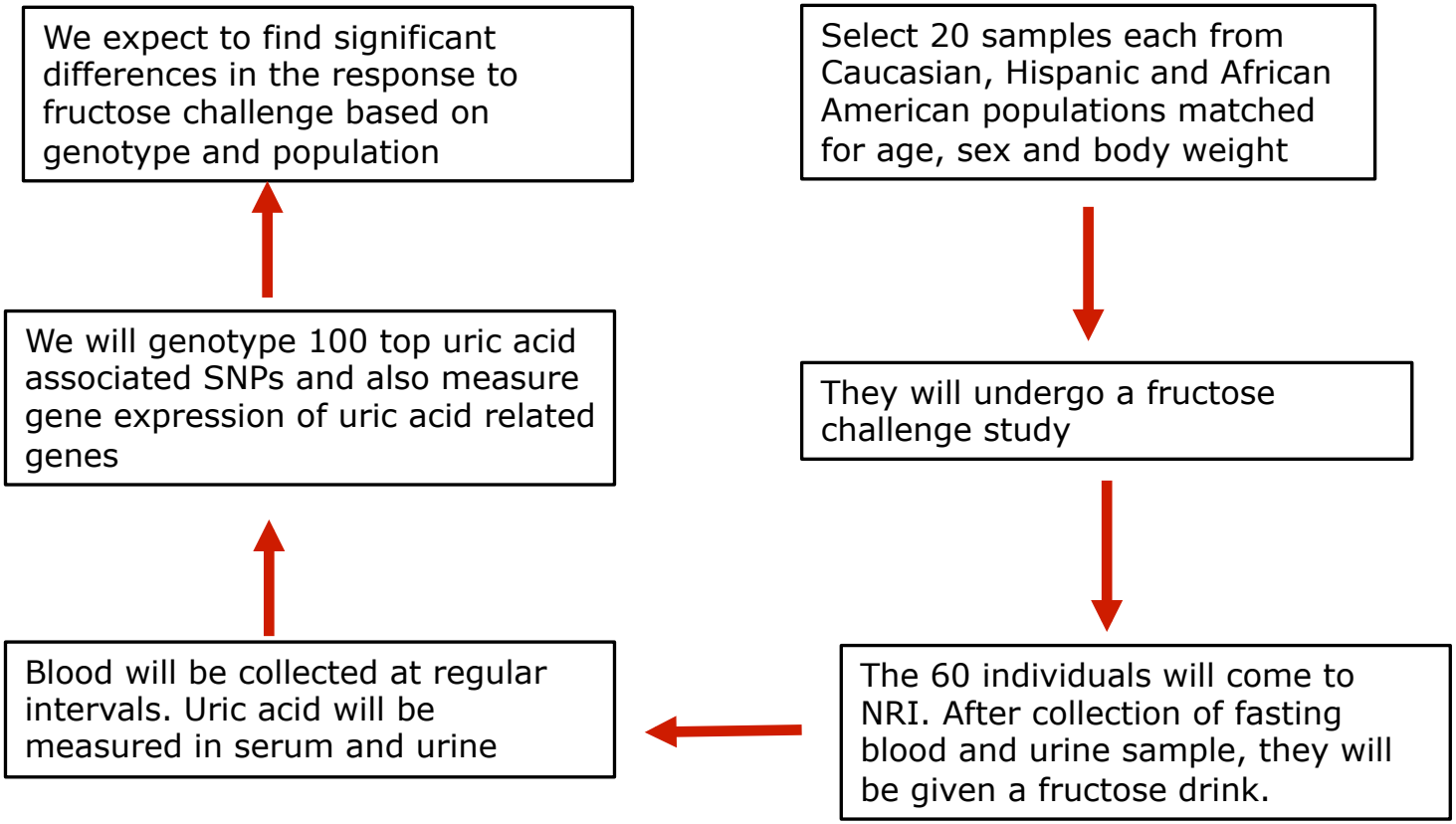
Effect of *ABCG2* genotype on serum uric acid concentrations during a fructose load



Dalbeth et al. Arthritis Research and Therapy. 2014; 16:R34

UNC NUTRITION RESEARCH INSTITUTE

Genotype- and population-specific effects of fructose on uric acid related genes



UNC NUTRITION RESEARCH INSTITUTE

- **UNC NRI faculty and staff**
- **NIH Grants**
  NIH R01 DK092238,
  NIDDK P01 DK056350
- **Participants of all studies**
- **Collaborators**
  Texas Biomedical Research Institute, San Antonio
  Baylor College of Medicine, Houston
  MURDOCK Study, Duke University

- **Voruganti Lab**

UNC NUTRITION RESEARCH INSTITUTE