The Women's Genome Health Study (WGHS): Recent progress and opportunities in the genetic epidemiology of aging

Daniel Chasman

Nutrition Omics symposium, HSPH

May 30, 2017

Women's Health Study (WHS)

- A trial of randomized aspirin (100 mg/alt day) and vitamin E (600 IU/alt day) v. placebo
- Primary prevention of CVD and cancer over 10 years ending in 2004
- Among 39,876 female healthcare professionals, 45+ yrs at baseline
- Now an NIH-funded observational cohort
- Julie Buring (PI) and I-Min Lee (co-PI)

- Not Nurses' Health Study, not Women's Health Initiative!

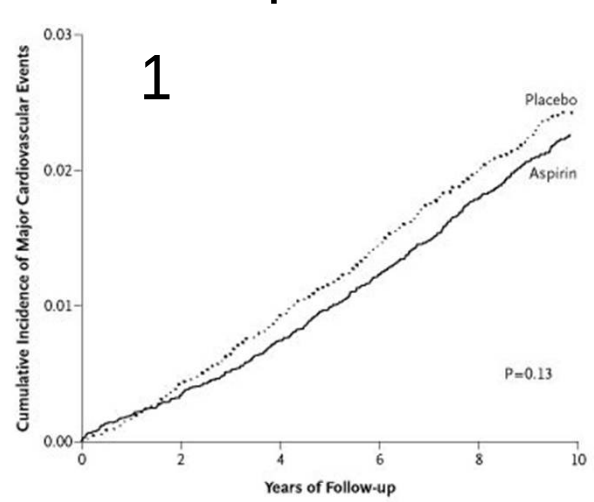
WHS results (2005)

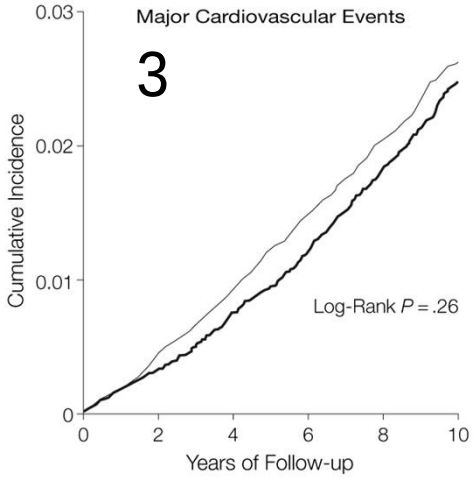
aspirin

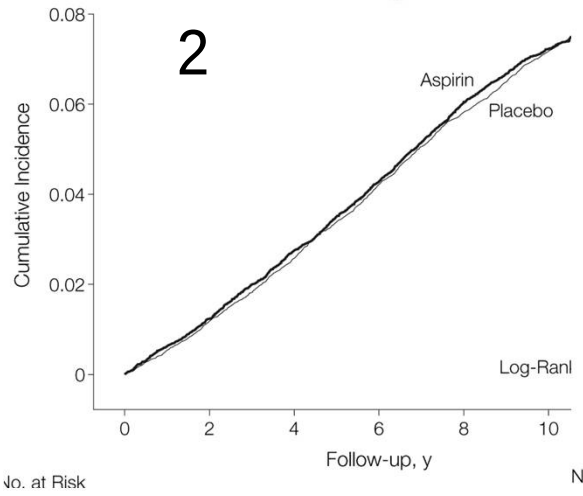
vitamin E

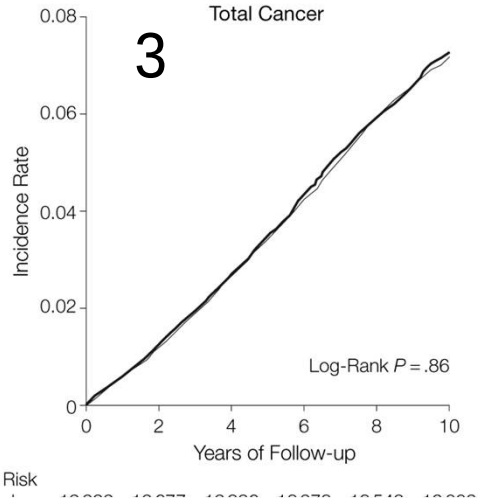
major CVD

total cancer









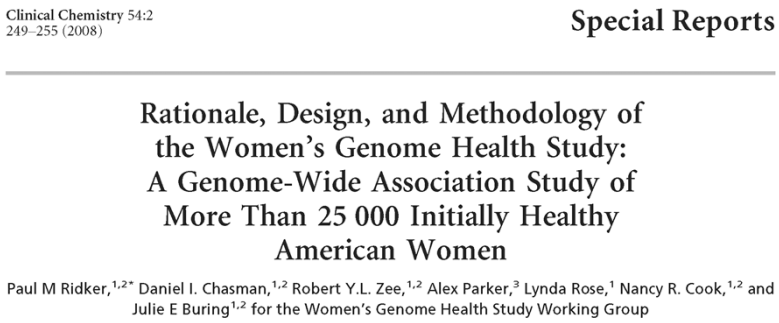
[1]Ridker et al. 2005 NEJM 352:1293; [2]Cook et al. 2005 JAMA 294:47; [3]Lee et al.2005 JAMA 294:56

WHS design

- Population recruited from
  - 1,757,247 invitations
  - 453,787 completed baseline questionnaire (194,659 willing, 65,169 eligible (14%) for placebo run-in
  - 39,876 selected after run-in, no overlap with NHS
- Questionnaires were at least annual during the trial
- Blood cohort among N=28,345 (71.1%)

Women's Genome Health Study (WGHS)

The WGHS is the WHS blood cohort augmented with whole genome genetic data



Clinical Chemistry 54:2
249–255 (2008)

Special Reports

Rationale, Design, and Methodology of the Women's Genome Health Study: A Genome-Wide Association Study of More Than 25 000 Initially Healthy American Women

Paul M Ridker,[1,2*] Daniel I. Chasman,[1,2] Robert Y.L. Zee,[1,2] Alex Parker,[3] Lynda Rose,[1] Nancy R. Cook,[1,2] and Julie E Buring[1,2] for the Women's Genome Health Study Working Group

Primary genetic data (2007-2009), Paul Ridker, PI

- Illumina HumanHap300 Duo "+" array (~360,000 SNPs)
- Custom content enriched for cardiovascular candidates and functional markers
- Sample size: whites: N=23,294, blacks=N=378
- Imputation 1000 Genomes

Secondary genetic data (2011), Daniel Chasman, PI

- Illumina Exome v. 1.1
- whites only: N=22,618

Sequencing through AFGen TOPMed project (2016-2017)

- 114 cases of "premature Atrial AF/TOPMed

WGHS baseline (cross sectional data)

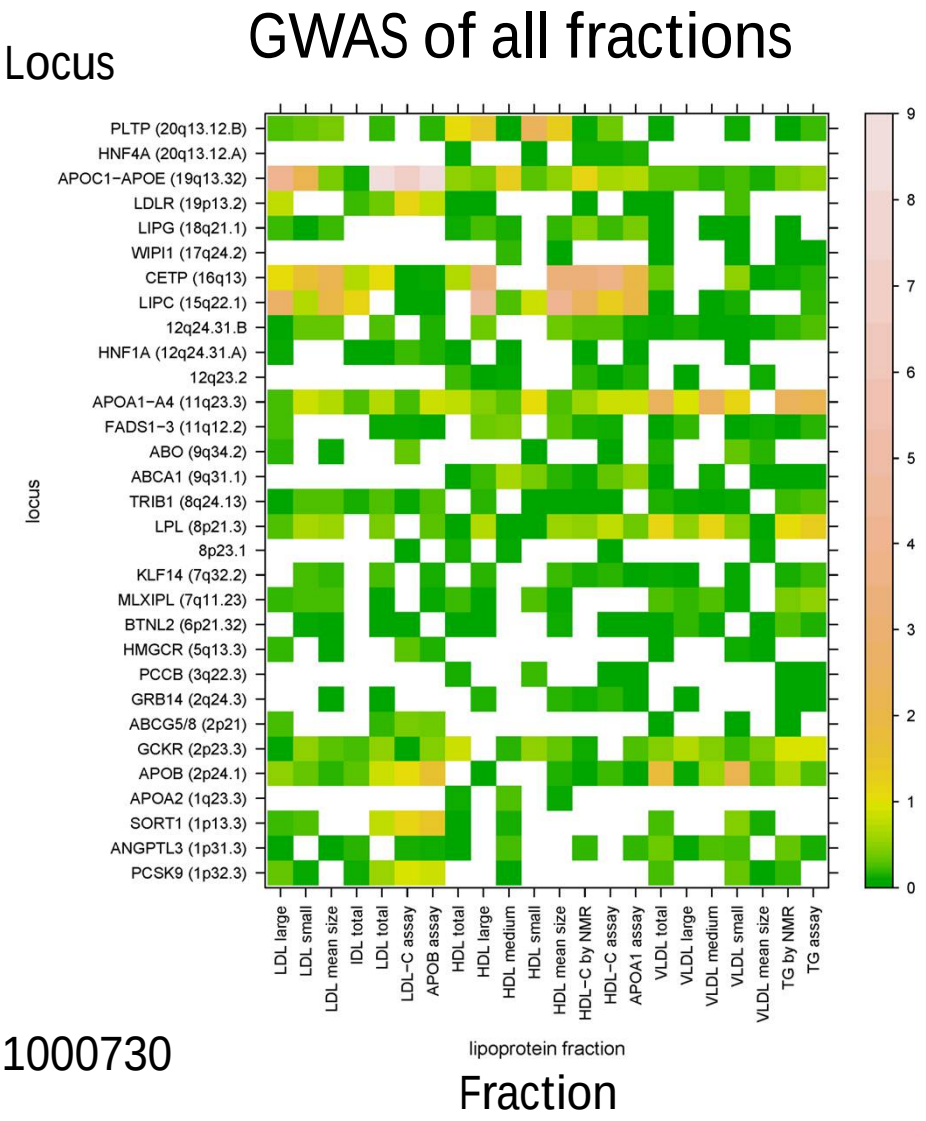
- Basic clinical profile
- Pre-existing conditions, reproductive aging and history, migraine, medications
- Diet QQ – same as NHS/HPFS
- Physical activity/alcohol/smoking
- Educational status
- Biomarkers in the blood sample (N=28,345)

  LDL-C, HDL-C, triglycerides, apoA1, apoB, Lp(a), C-reactive protein, ICAM-1, fibrinogen, homocysteine, creatinine, HbA1c

  GWAS on all of these biomarkers

Lipoprotein profile from NMR

- Lipoprotein sub-fractions, e.g. LDL, HDL, and VLDL particle concentration and size (NMR)
- entire sample (Samia Mora)



Chasman et al. PLoS Genetics 2009 5(11):e1000730

Additional, investigator-initiated, cross-sectional data during f/u

By self-report

- Perceived stress (Michele Albert)
- Physical activity accelerometer (I-Min Lee)
- Hearing/fibroids (Cynthia Morton)
- Endometriosis (Cynthia Morton, Kathy Rexrode)
- Handedness (Guillaume Pare)
- Osteoporosis

Prospectively ascertained endpoints

- Primary endpoints, CVD and cancer, are confirmed by physician review of medical records according to WHO standards
- Some other endpoints also adjudicated by investigator-initiated research

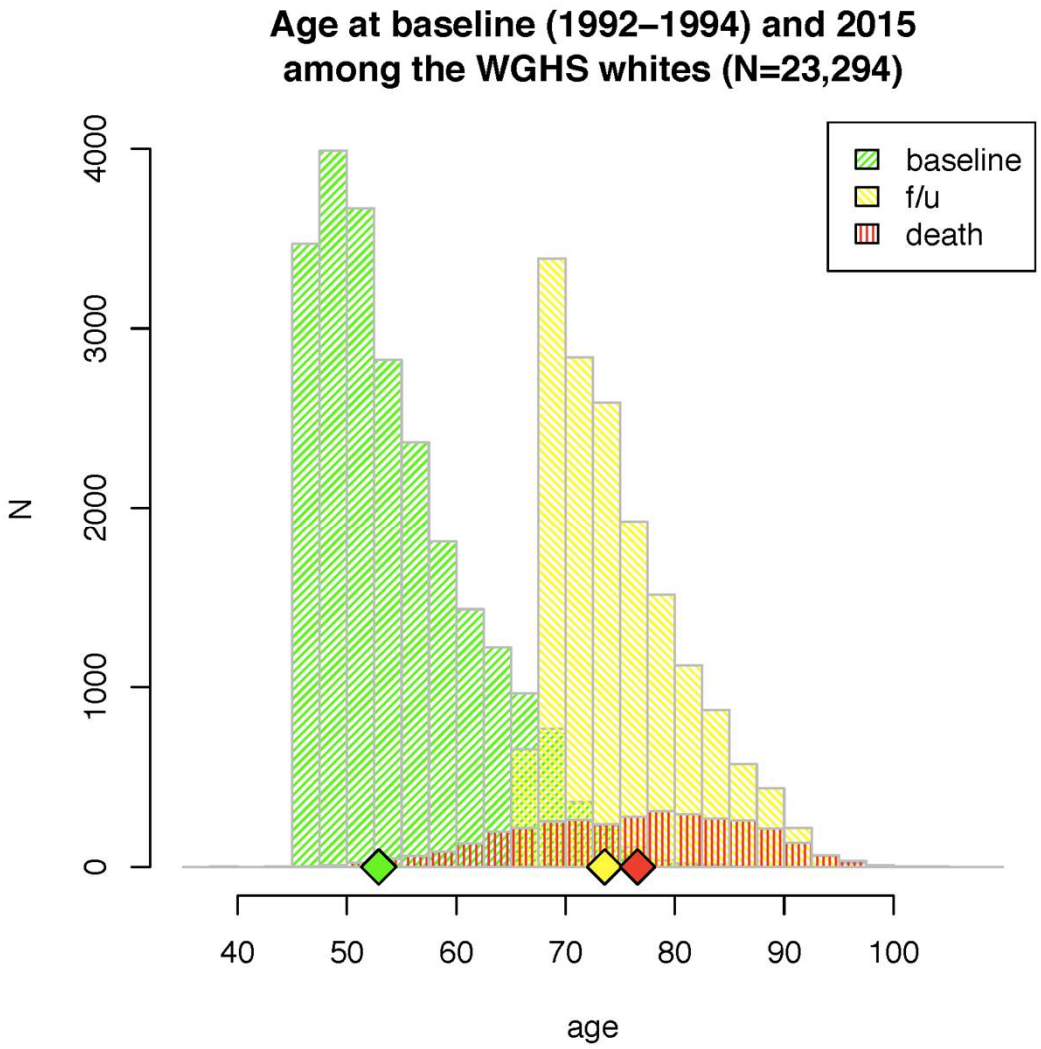
CVD events (2015)	
Event	N
Major CVD	1178
MI	451
Isch. stroke	552
CHD	1122
Total CVD	1782

Cancer events (2015)	
Event	N
Total	3672
breast	1524
CRC	309
lung	294
pancreas	60

Other events
Incident diabetes (N=1927)
Incident Afib. (Albert, N=959)
Incident hypertension
VTE
Longitudinal BMI (obesity)
AMD
Cognitive decline (Grodstein)

WGHS age distribution baseline (1992-1994) and 2015

- In 2017, f/u for 25 yrs
- Mortality near 100% through National Death Index
- >88% of surviving participants return questionnaires (as of ~2013)



Emphasis areas of genetic research

- Cardiovascular disease
  CVD risk factors, lipids/lipoproteins, inflammation measures, BP, incident disease
- AFib
- Diabetes
- AMD
- Kidney function
- Women's health
  - Menarche, menopuase, breast cancer, migraine, fibroids
- Migraine (more prevalent among women)
- Nutrition genetics

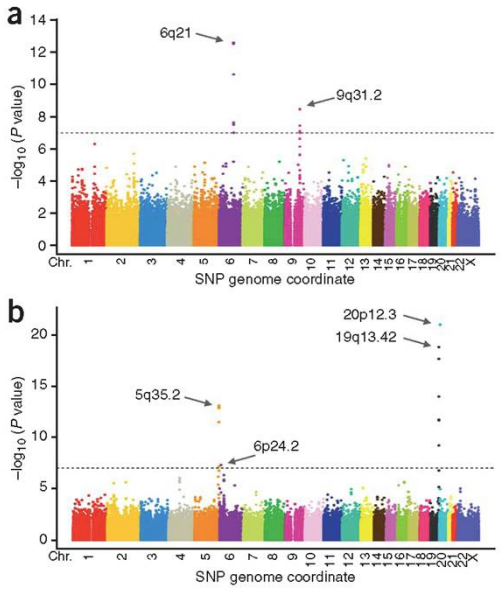
Participation in major GWAS consortia

- CHARGE (Heart and Aging)
  Working groups: AFGen, ReproGen, Inflammation, T2D, Neuro, Mitochondria, BP, INVENT, Nutrition, Gxlifestyle, Hematology
- GIANT, including GxE (Anthropometry)
- Global Lipids (circulating lipids)
- ICBP (blood pressure)
- IHGC (headache)
- DietGen (w/NHS/HPFS) (diet)
- LifeGen (longevity)
- CARDIoGRAM (CAD)
- BPC3 (breast cancer)
- GEFOS (bone)

Results: Reproductive aging (w/NHS)

Genome-wide association studies identify loci associated with age at menarche and age at natural menopause

Chunyan He[1,2], Peter Kraft[1,2], Constance Chen[1,2], Julie E Buring[2–4], Guillaume Paré[3–5], Susan E Hankinson[2,6], Stephen J Chanock[7], Paul M Ridker[2–5], David J Hunter[1,2,6–8] & Daniel I Chasman[3–5]



Now with ReproGen

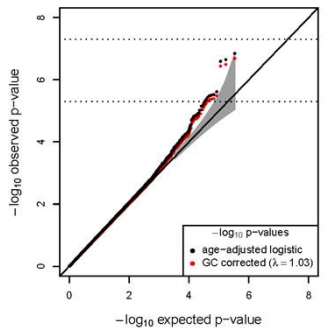
- menarche
  - 106 loci (development, parent-of origin, interaction w/age!)
- natural menopause
  - 44 loci (DNA repair)

He et al. Nature Genetics 2009 41:724

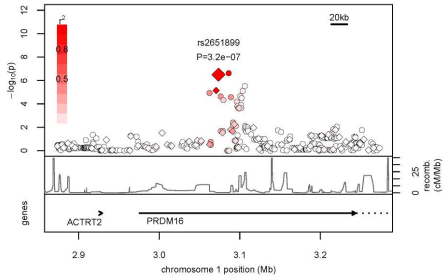
Results: Migraine

2011 (WGHS alone)

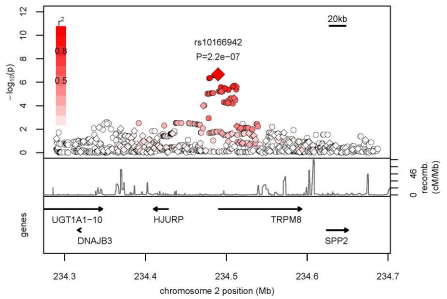
Genome-wide QQ plot



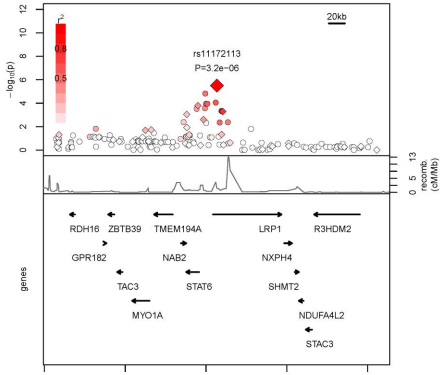
PRDM16



TRPM8



LRP1



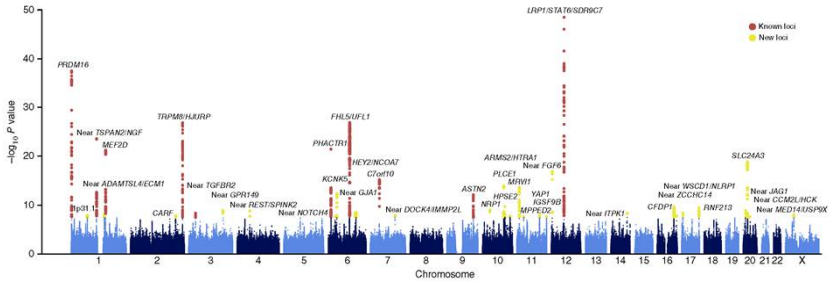
Now (2016) with the International Headache Genomics Consortium (IHGC)

38 loci (strong emphasis on vascular development)

Nature Genetics 2016 48:856

Meta-analysis of 375,000 individuals identifies 38 susceptibility loci for migraine

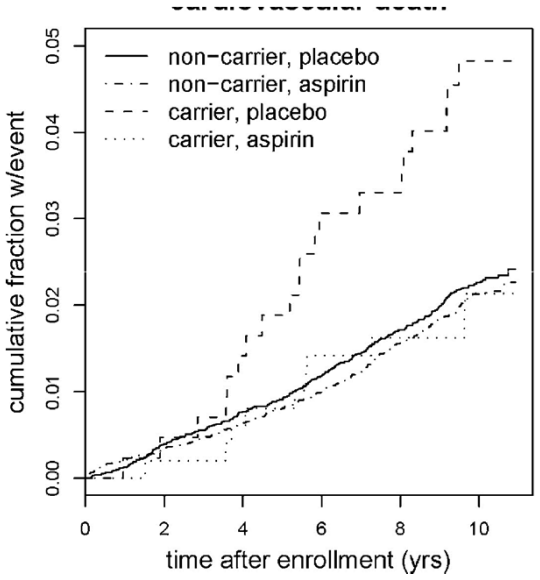
Padhraig Gormley[1–4,81], Verneri Anttila[2,3,5,81], Bendik S Winsvold[6–8], Priit Palta[9], Tonu Esko[2,10,11], Tune H Pers[2,11–13], Kai-How Farh[2,5,14], Ester Cuenca-Leon[1–3,15], Mikko Muona[9,16–18], Nicholas A Furlotte[19], Tobias Kurth[20,21], Andres Ingason[22], George McMahon[23], Lannie Ligthart[24], Gisela M Terwindt[25], Mikko Kallela[26],



Results: G x aspirin interaction during WHS trial

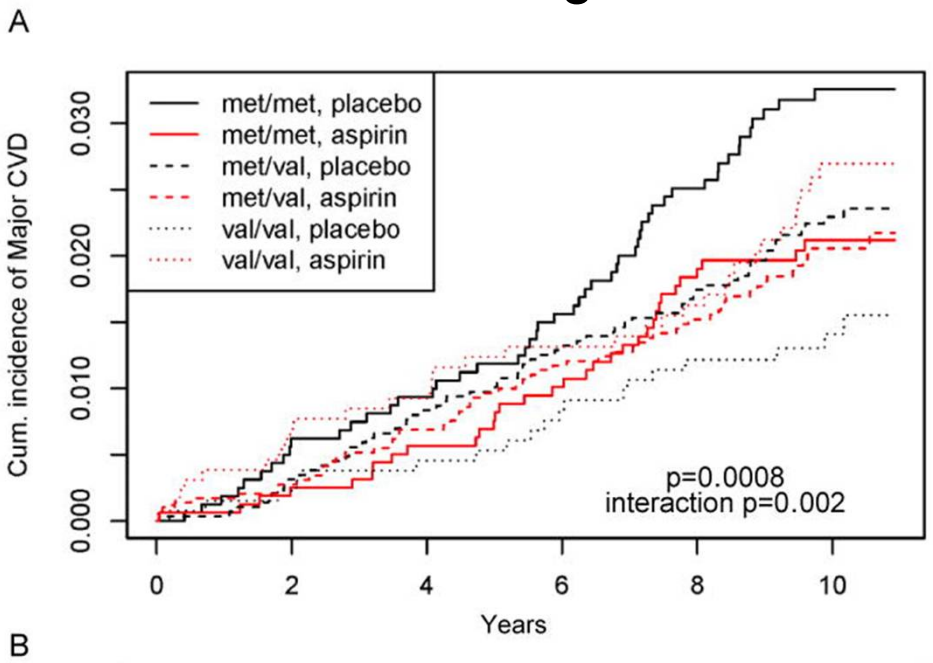
Aspirin effects on incident CVD stratified by genotype

LPA gene



Atherosclerosis 2009 203:271

COMT gene



ATVB 2014 34:2160 (Kathryn Hall)

Results: Genetic risk score (GRS) v. clinical prediction

GRS: CVD + risk factors

GRS: CVD

FamHx MI

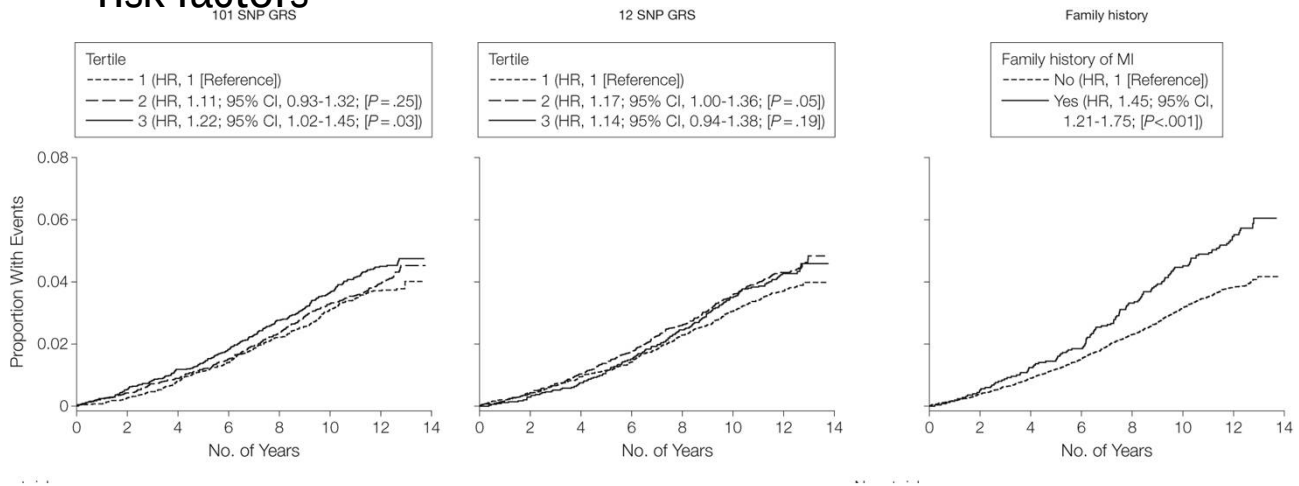


Table 2. Discrimination and Reclassification After Addition of Genetic Risk Score (GRS) or Family History of Cardiovascular Disease (CVD) to Base Model

		101 SNP GRS[a]				12 SNP GRS[b]				Family History of Premature MI[c]			
		Discrimination		Reclassification		Discrimination		Reclassification		Discrimination		Reclassification	
	Base Model C Index	C Index	*P* Value[d]	NRI	*P* Value[e]	C Index	*P* Value[d]	NRI	*P* Value[e]	C Index	*P* Value[d]	NRI	*P* Value[e]
Age	0.701	0.704	.14	1.2	.13	0.705	.01	0.6	.52	0.709	.01	3.1	.02
Covariates													
ATP III[f]	0.793	0.793	.92	0.5	.24	0.794	.12	0.5	.59	0.796	.06	1.4	.28
Reynolds[g]	0.796	0.796	.84	0.4	.21	0.796	.12	0.8	.36	NA	NA	NA	NA

Abbreviations: ATP III, Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults; MI, myocardial infarction; NA, data not applicable; NRI, net re-

JAMA 2010 303:631 (Nina Paynter)

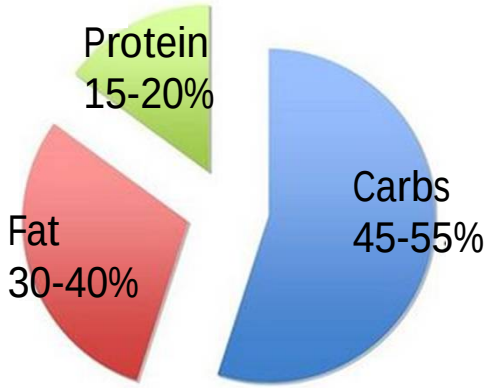
Results: BP by self report

- Baseline BP was self-reported by categories SBP (10), DBP (7)
- Reluctance in the genetics community to use these measures
- With Framingham HS investigators* assessed BP loci the WGHS at loci from the CHARGE (29,136) and GlobalBPGen (N=34,433) consortia GWAS (excluding WGHS)
- High concordance and formal replication
  - Among 43 candidate SNPs 13/18 primary, 3/13 secondary, SNPs, and 4/12 eSNPs met $p < 0.05/43$
- Meta-analysis
  - 1 new locus (BLK-GATA4) and 1 new genome-wide significant SNP at CASZ1 in Europeans

*Jennifer Ho, Andrew Johnson, & Dan Levy. Ho et al. J. Hypertension 2011 29:62.

Macronutrients (Audrey Chu)

- Carbohydrates, protein and fat
- Absolute intakes are highly variable
- Average intake is more consistent across populations when calculated as percentage of total caloric intake



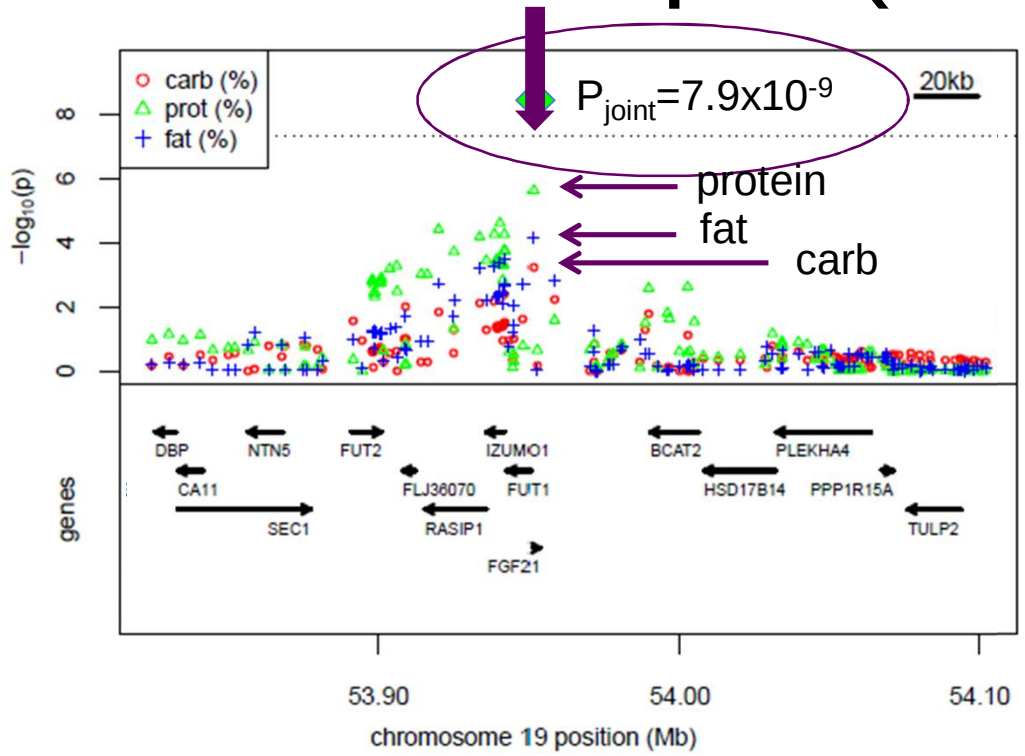
- Heritability of macronutrient intakes ~ 11-65%

Mann 2007; Rankinen 2006

Methods

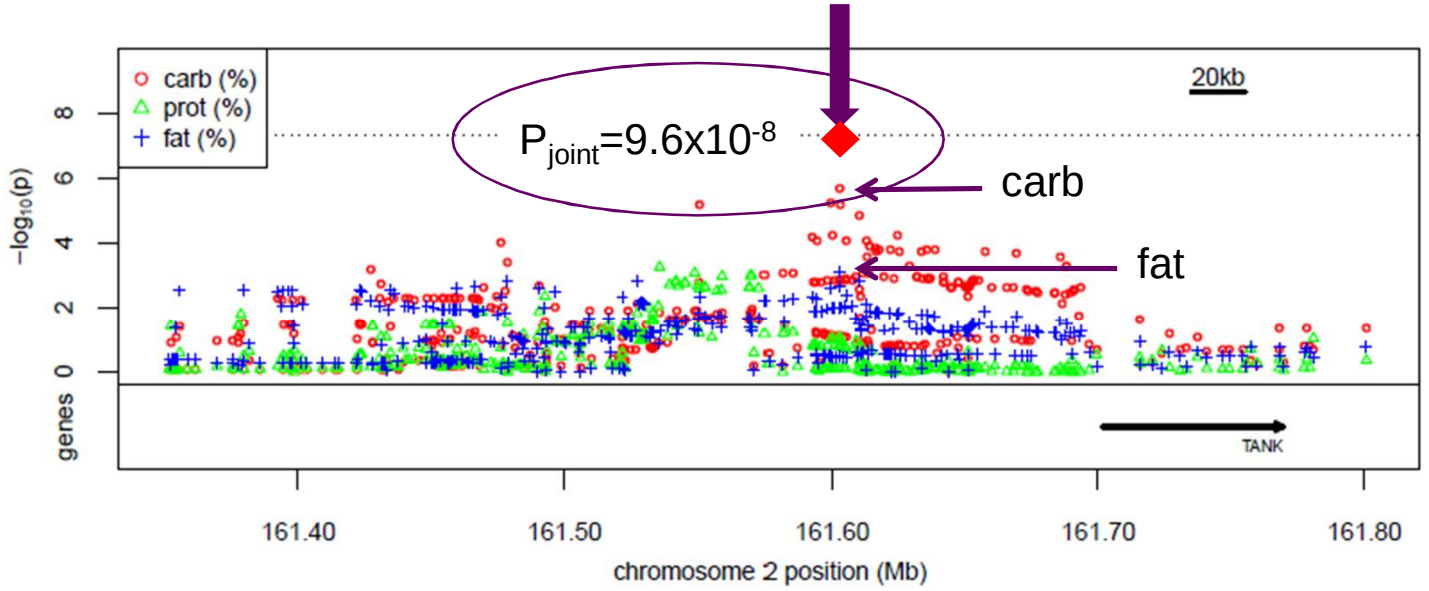
- DietGen Consortium (N=33,338)
  - Health Professional's Follow-up Study (HPFS)
  - Nurses' Health Study (NHS)
  - Women's Genome Health Study (WGHS)
- Genotyping – Affymetrix- or Illumina-based genotyping panels imputed to 2.6 million SNPs (HapMap release 22)
- Phenotyping – Percentage of total energy intake from carbohydrate, protein and fat derived by Willett FFQ
- Genetic association analysis
  - Each SNP was adjusted for age, location and population stratification in an additive model, +/- BMI
  - Fixed-effects meta-analysis, MAF ≥ 0.05

Chromosome 19q13 (*FGF21*)



- rs838133: synonymous SNP in *FGF21* gene
- DietGen: β -0.12% protein, 0.23% carbs, -0.21% fat per minor allele
- CHARGE association for protein: β -0.10% ($P=7.3x10^{-4}$)

Chromosome 2p24 (*TANK*)



- rs197273: intergenic SNP near *TANK*
- DietGen: β 0.31% carbs, -0.17% fat per minor allele
- CHARGE association with carbohydrates: β 0.17% (P=2.9x10^{-3})

Biological Context

- *FGF21* (fibroblast growth factor-21)
  - Stimulates glucose uptake in adipocytes
  - Expression can be induced through fasting and feeding in rodents
  - Administration increases energy expenditure in mice
  - Associated with T2D and obesity
- *TANK* (TRAF family member-associated activator)
  - Binds to cytokines in inflammation and immunity
  - Not previously linked to dietary intake

Kharitonenkov 2005; Uebanso 2011; Sarruf 2010; Chen 2008; Tan 2011

AlcGen paper*

- GWAS of alcohol consumption 70,460 people, 30 cohorts with replication f/u
- Only *KLB* locus genome-wide significant with f/u sample
- *TANK* SNP genome-wide significant initially but not combination with f/u sample
- *KLB* gene (β Klotho) has reciprocal effects to FGF21

*PNAS 2016 113: 14372

Heritability Results

Estimated proportion of variance in macronutrient intake explained by common tag-SNPs in the WGHS (GCTA)

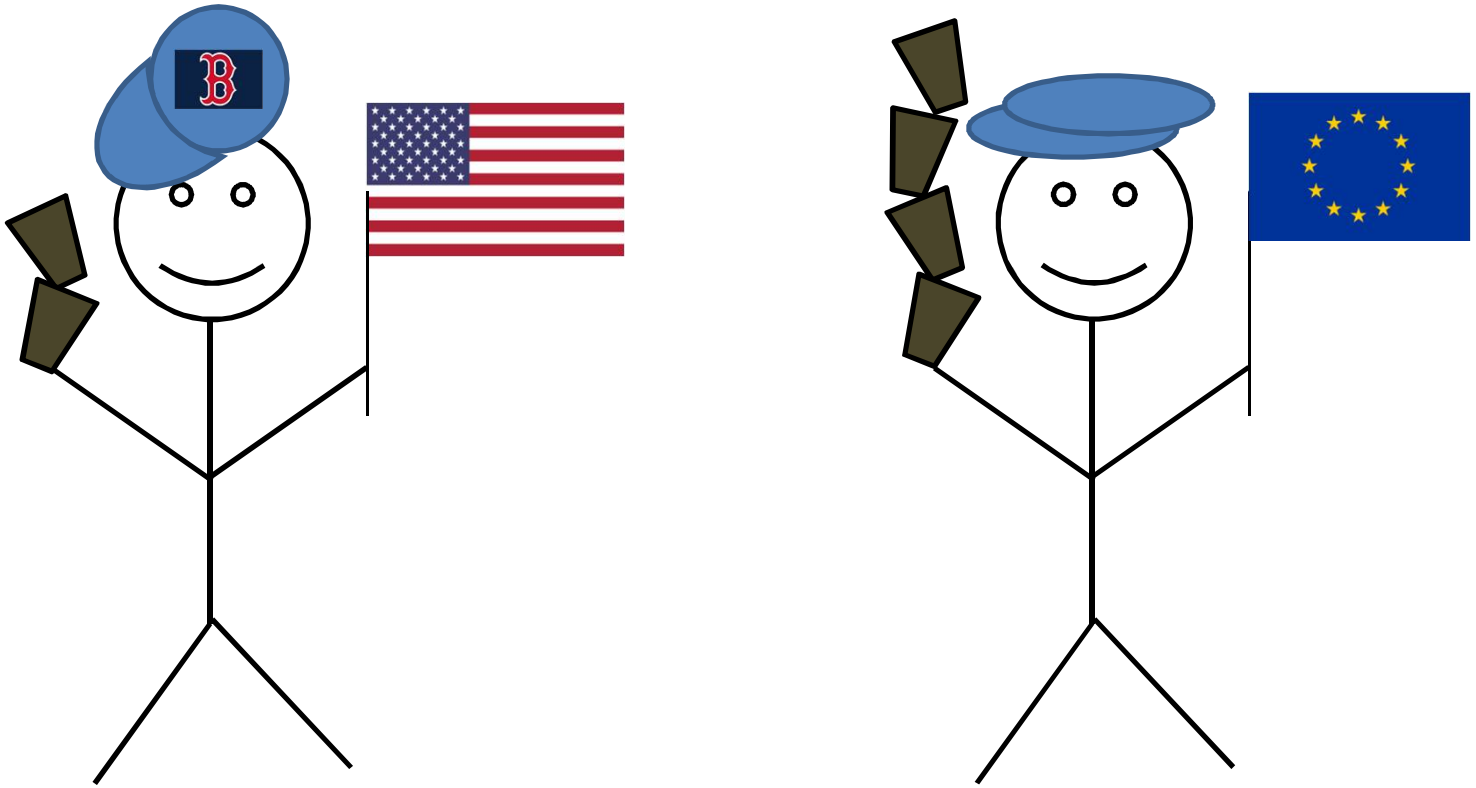
	Genome-wide			*FGF21*
	h^2	SE	P	h^2
Carbohydrates	6.2%	2.1%	2.1×10^{-4}	0.052%
Protein	8.0%	2.2%	1.4×10^{-5}	0.062%
Fat	6.4%	2.1%	3.2×10^{-4}	0.057%

Adjusted for age, location, population substructure

- Additional variants to be identified
- Heritability underestimated due to use of tagging SNPs to estimate relatedness

Yang 2011

What determines coffee drinking habits?



Some influences are cultural

Coffee and Caffeine Genetics Consortium (Marilyn Cornelis*)

- Populations

  >100,000 individuals from 48 studies in the US, Europe, and Pakistan

- Coffee consumption measure

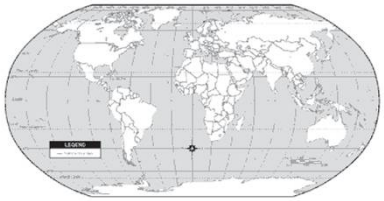
  cups/day among coffee drinkers by self-report

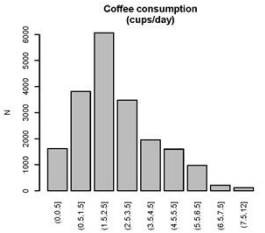
- Genetic data

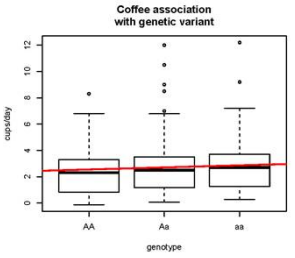
  HapMap-based analysis (2.6 million SNPs)

- Statistical test

  Test association between self-reported coffee consumption and genetic variation

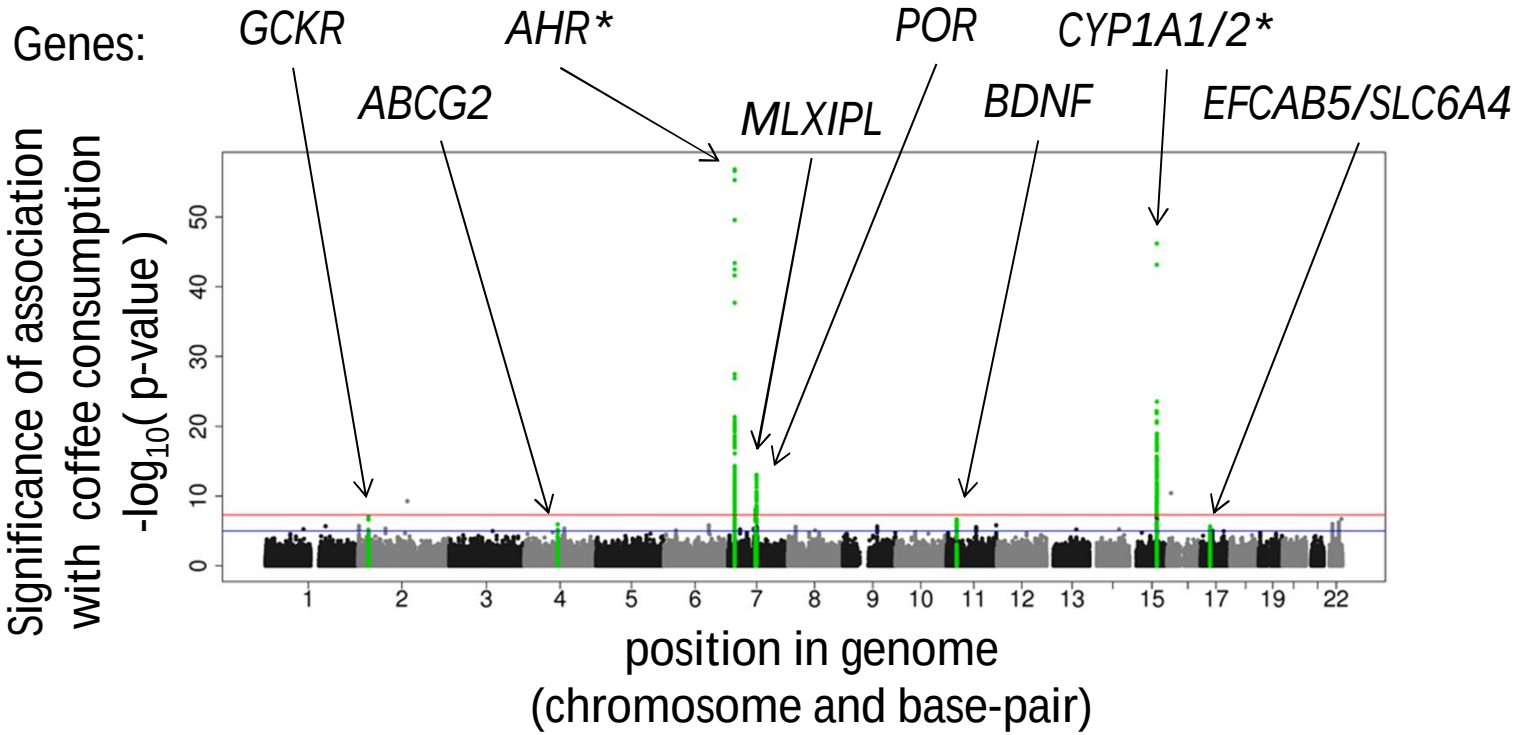






*Formerly HSPH, now at Northwestern

GWAS findings: 8 genes, 6 novel

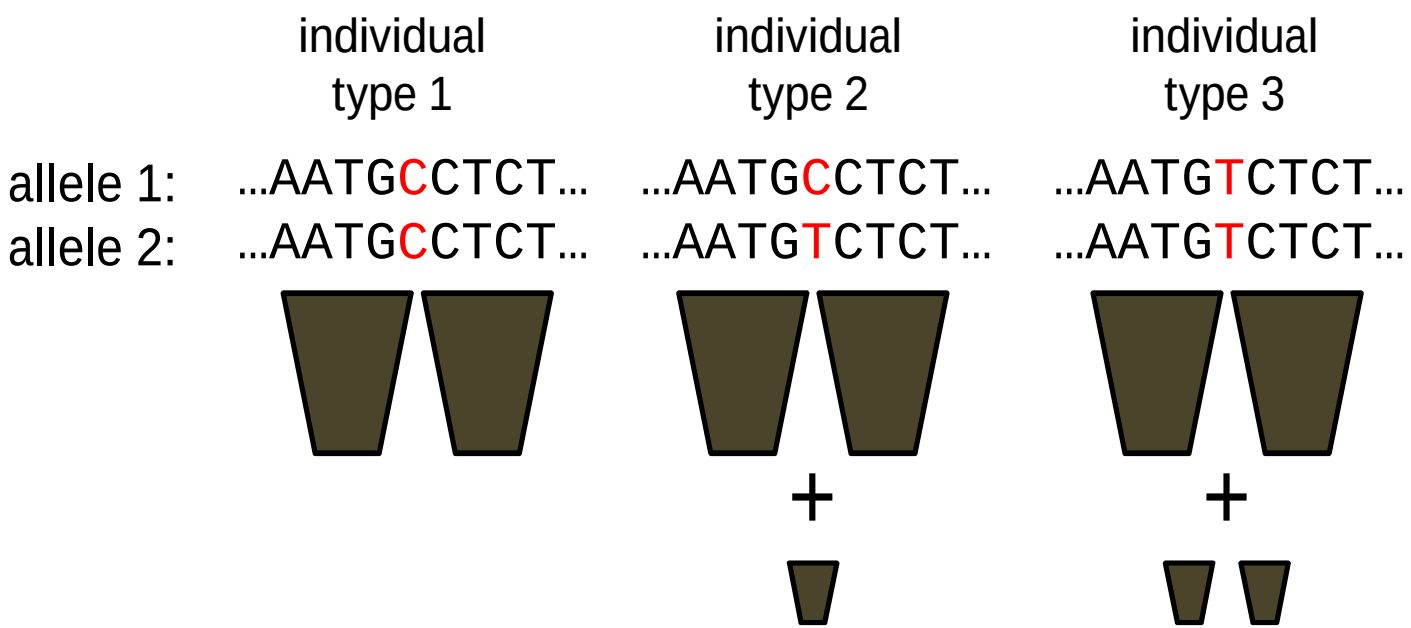


*Identified previously in genetic analysis of habitual caffeine consumption

How big are the effects?

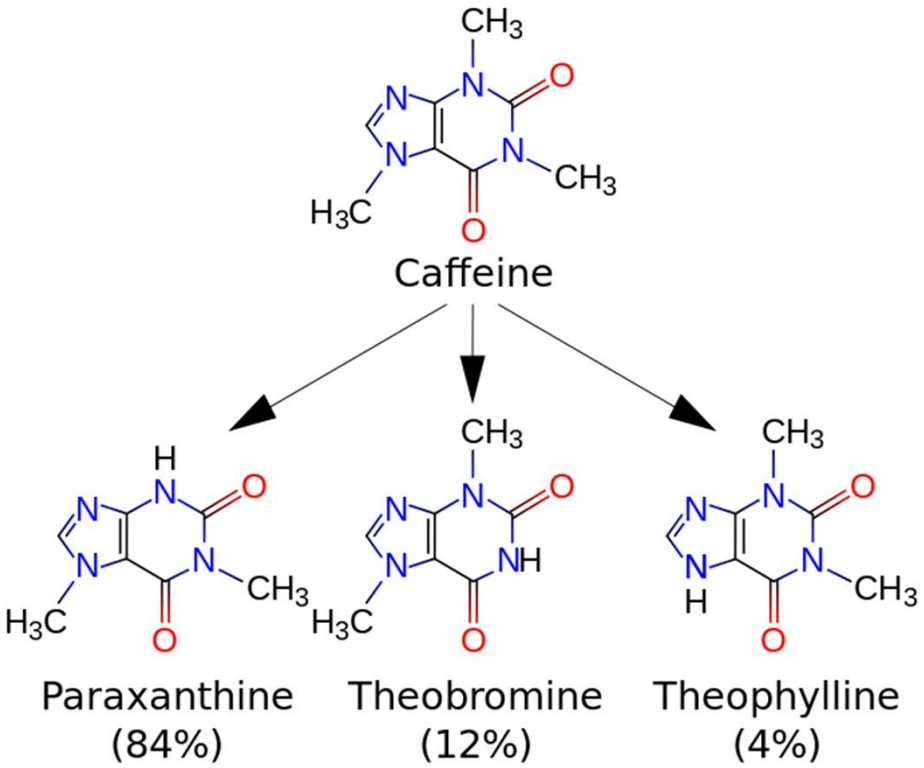
Average coffee consumed for different genotypes at the CYP1A1/2 gene(s):

0.14 additional cups/day per inherited "T"-allele compared to "C" allele



Range at other genes: 0.04 - 0.12 cups/day

Four genes involved in caffeine (xenobiotic) metabolism



Metabolic functions

- Degrading caffeine

  *CYP1A1/2** – cytochrome p450

  POR - P450 (cytochrome) oxidoreductase
- Regulating p450 enzymes

  *AHR** - aryl hydrocarbon receptor
- Excreting caffeine and its metabolites

  ABCG2 - ATP-binding cassette G2

*Identified previously in genetic analysis of habitual caffeine consumption

Remaining genes: two functions

A. Genes involved in behavior, psychiatric traits, and mood

BDNF – "brain-derived neurotrophic factor." Associated with smoking initiation, obesity/body mass index (eating) and modulates activity of serotonin, dopamine, and glutamate

SLC6A4 – "solute carrier family 6 member A4." Transports serotonin into neurons.

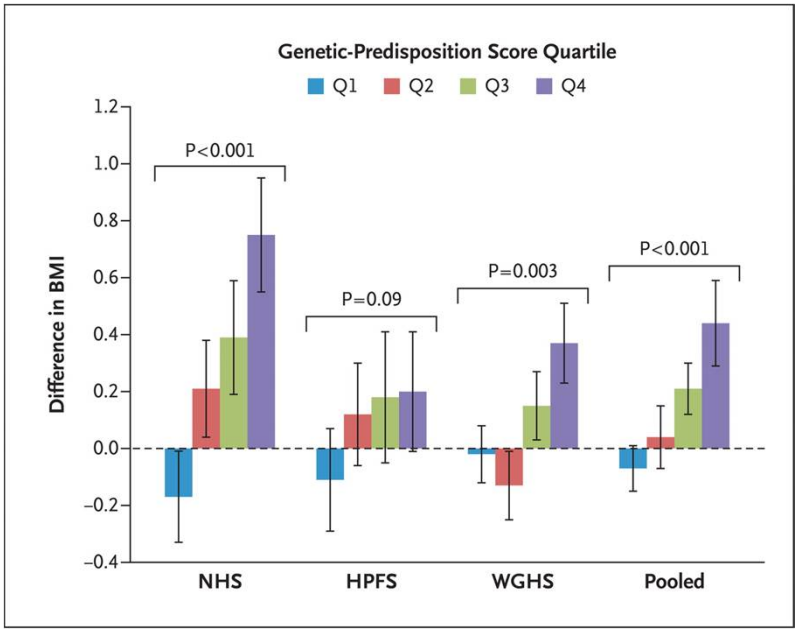
B. Genes involved in glucose metabolism, a possible link to coffee's protective effect on diabetes.

GCKR – "glucokinase regulatory protein." Associated with plasma glucose, insulin levels, triglycerides, other lipids, inflammation markers, etc.

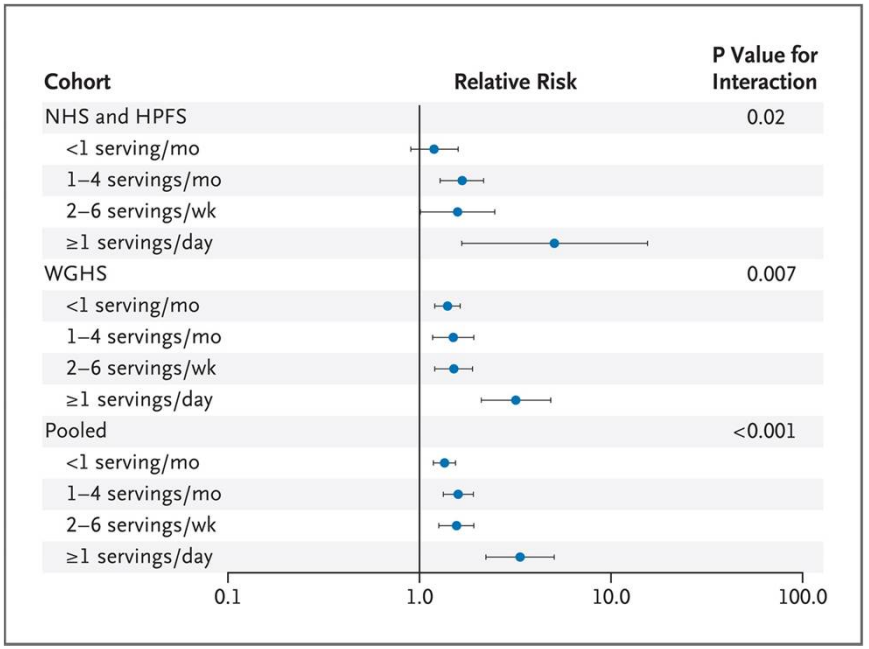
MXLIPL – "MLX interacting protein-like." Regulates pathways that influence triglyceride levels in response to plasma glucose.

Genetic interaction with sugar-sweetened beverages (Qibin Qi, Lu Qi)

Cross-sectional BMI increase per SSB serving



Incident obesity (prospective) WGHS replication w/~6 yrs f/u



NEJM 2012 367:1378

Recap

- The Women's Genome Health Study (WGHS) is a large prospective cohort for whole genome genetic analysis among women
- Primary endpoints in CVD and cancer
- Follow-up is ongoing, now into 25th year
- Mean age from ~52 to ~73 (2015)
- Rich data source at baseline including detailed clinical profile, blood-based biomarkers, dietary habits

WGHS contributors

Investigators from DPM
Paul Ridker (WGHS PI)
Daniel Chasman (WGHS co-PI)
Julie Buring (WHS PI)
I-Min Lee (WHS co-PI)
Christine Albert
Michele Albert
Nancy Cook
Brendan Everett
Kathryn Hall
Tobias Kurth
Samia Mora
Nina Paynter
Aruna Pradhan
Kathy Rexrode
Debra Schaumberg
Robert Zee

DPM fellows/scientists
Shafqat Ahmad
Tunde Akinkuolie
Audrey Chu
David Conen
Christina Ellervik
Franco Giulianini
Amit Khera
MV Moorthy
Patrick Lawler
Guillaume Pare
Lynda Rose
Marty Vandenburgh
Yau Hua Yu

NHS/HPFS collaborators
Marilyn Cornelis
Ming Ding
Fran Grodstein
Chunyan He
Frank Hu
Tao Huang
David Hunter
Peter Kraft
Po-Ru Loh
Alkes Price
Lu Qi
Qibin Qi
Bjarni Vilhjalmsson
Tiange Wang
Noah Zaitien

WHS/WGHS participants

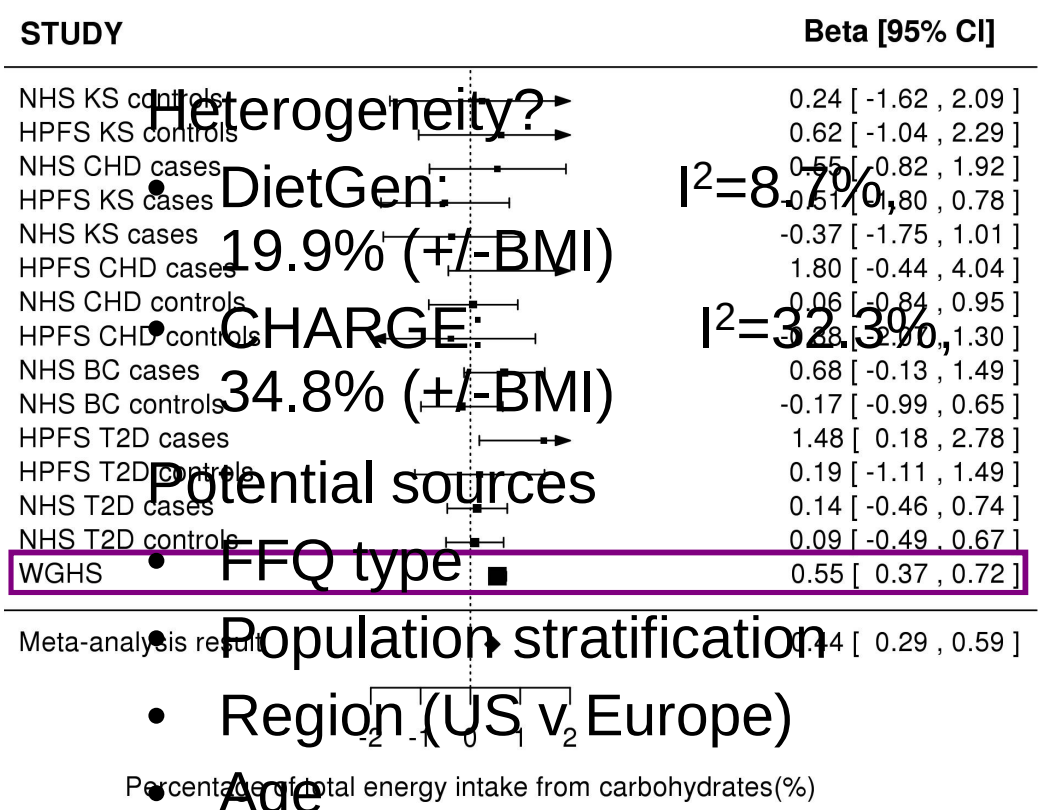
Funding
NIH
NHLBI (HL043851, HL080467)
NCI (CA047988, UM1CA182913
Amgen
Donald W. Reynolds Found.

Chromosome 16q12 (*FTO* locus)

Association at *FTO* not in LD with BMI associated SNP

Replication

- rs10163409 for carbohydrate intake, P_{CHARGE}=0.34
- Association driven by WGHS in DietGen, β(se): 0.55(0.09), $P_{WGHS}=5.7x10^{-9}$
- Without WGHS, β(se): 0.20(0.14), P_{noWGHS}=0.15



- Heterogeneity?
  - DietGen: I^2=8.7%, 19.9% (+/-BMI)
  - CHARGE: I^2=32.3%, 34.8% (+/-BMI)
- Potential sources
  - FFQ type
  - Population stratification
    - Region (US v. Europe)
    - Age
    - Menopause status
    - Sex