Novel Findings in Genomics and Metabolomics in the ARIC study

Eric Boerwinkle
Boston, MA
May, 2017

UTHealth
The University of Texas
Health Science Center at Houston

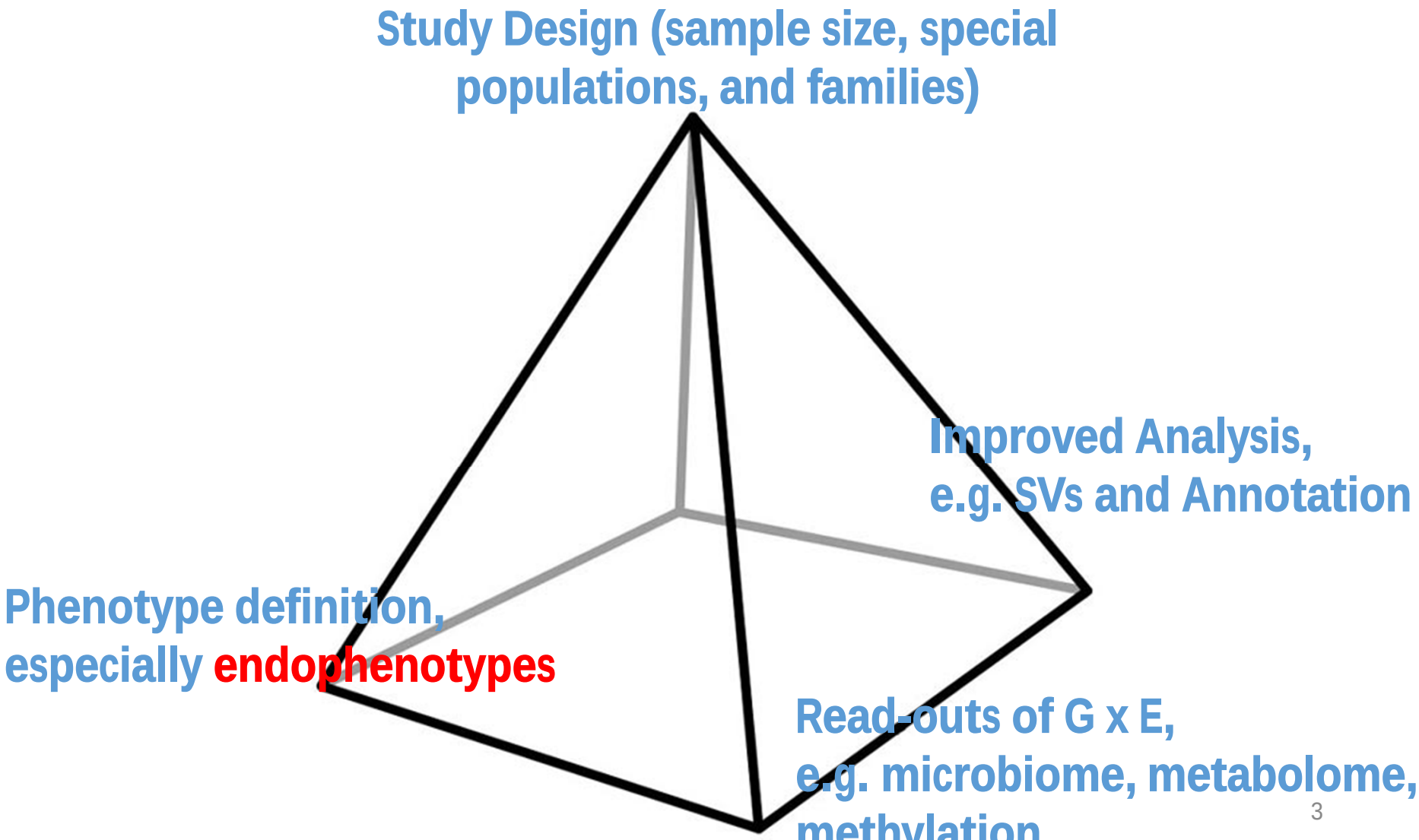
HGSC
HUMAN GENOME SEQUENCING CENTER

Goals of Genetic Studies
(of the Metabolome)

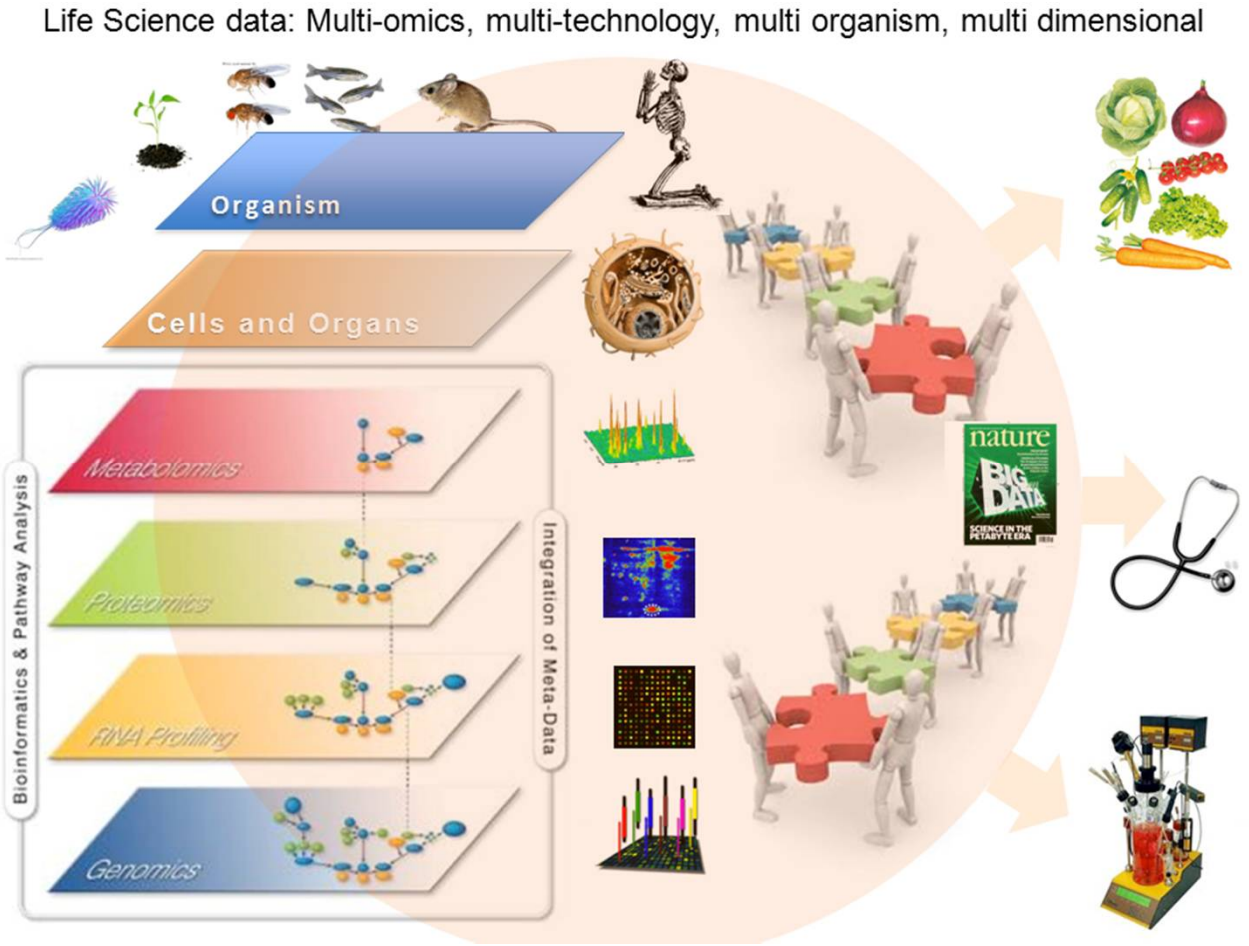
- Genes being novel predictors of disease
- Predictors vs Biomarkers and the principal of Mendelian Randomization
- Biology of the human metabolome
- Drug Target Discovery
- Gene x Environment Interaction

Maximizing Opportunity for Discovery

↑ Power, while controlling costs



Multi-Omics Integration

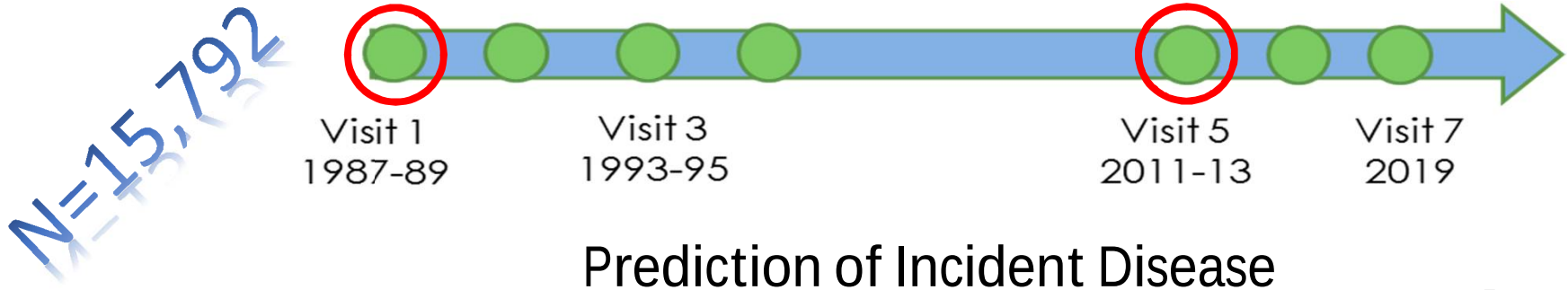


Achieving this vision, requires delivering large amounts of high quality data to the community in a timely manner.

The Atherosclerosis Risk in Communities (ARIC) Study

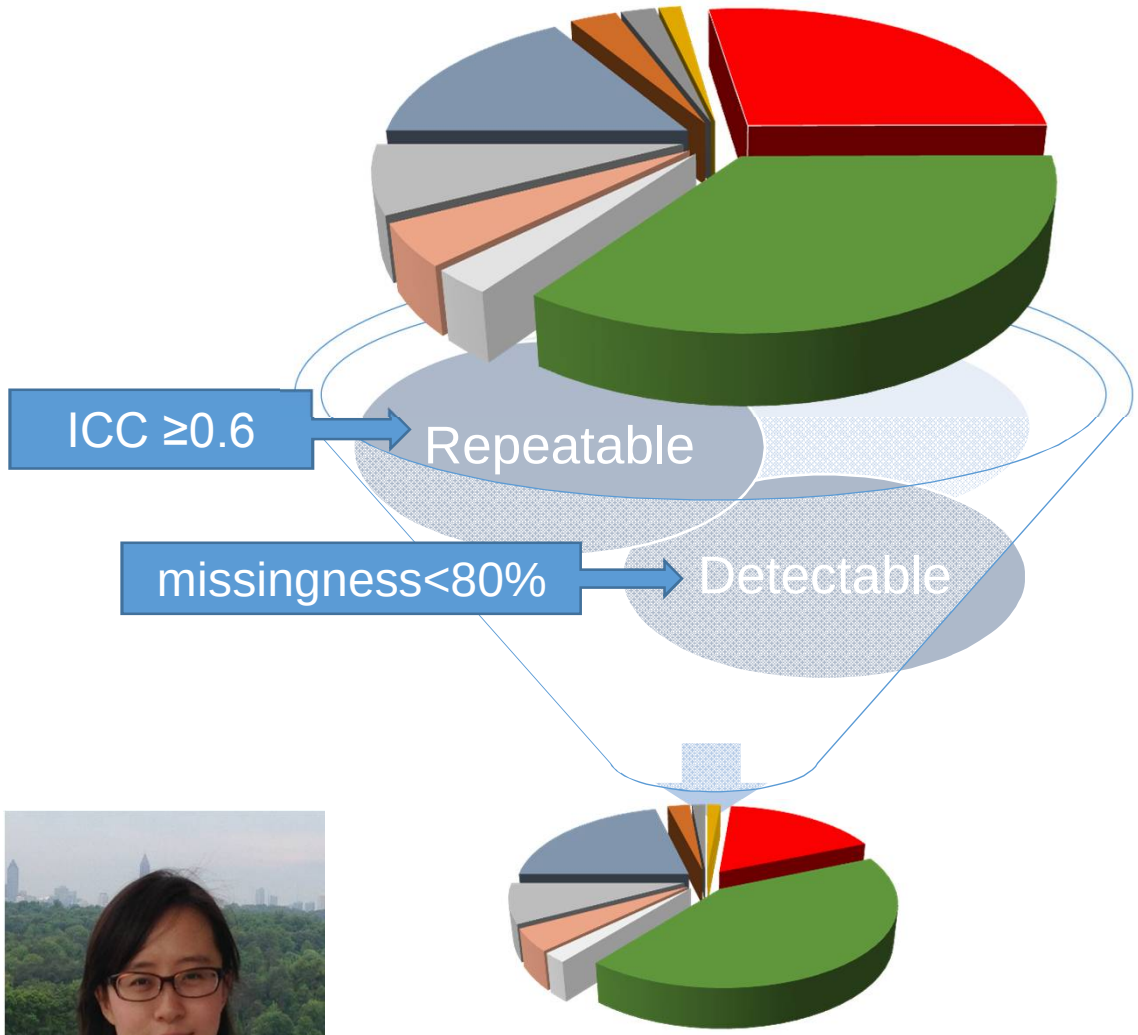
- 1,679 African-Americans (AAs) among the ARIC study
- 1,458 ARIC European Americans (EAs)

Having metabolomics data



Prediction of Incident Disease

Metabolomics in the ARIC study



METABOLON
SOLUTIONS BEYOND THE DATA...™

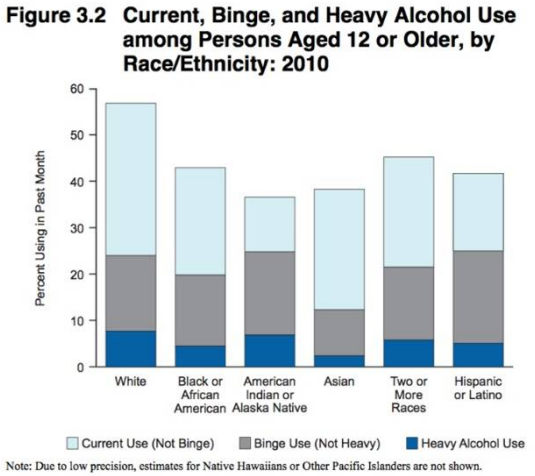
COMPOUND TYPE	COUNT
Named Metabolites	361
Unnamed Metabolites	241
Total Number of Measured Metabolites	**602**

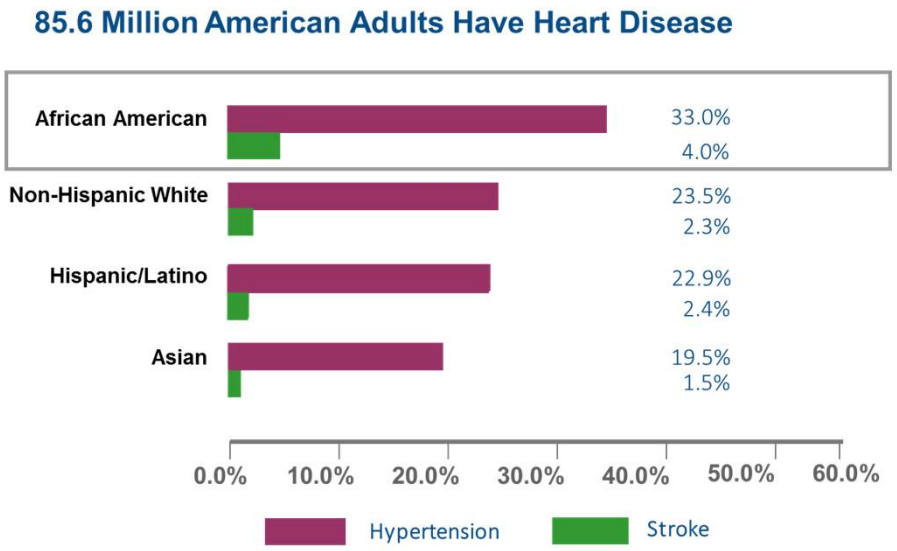
- Amino acid
- Carbohydrate
- Cofactors and vitamins
- Energy
- Lipid
- No Super Pathway
- Nucleotide
- Peptide
- Xenobiotics

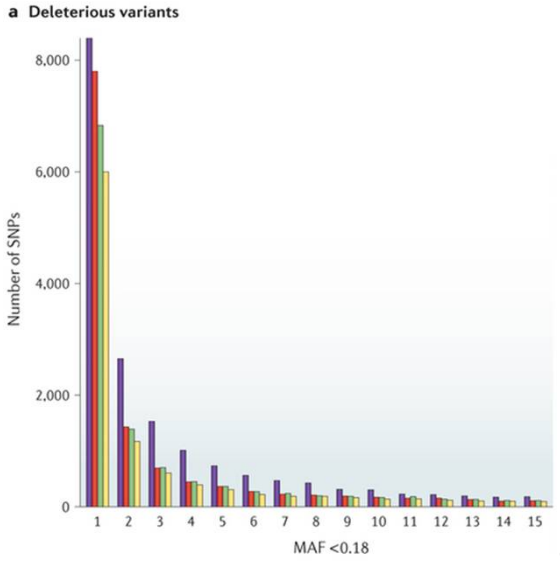
COMPOUND TYPE	COUNT
Named Metabolites	118
Unnamed Metabolites	86
Total Number	**204**

Why Multi-Ethnic Studies?

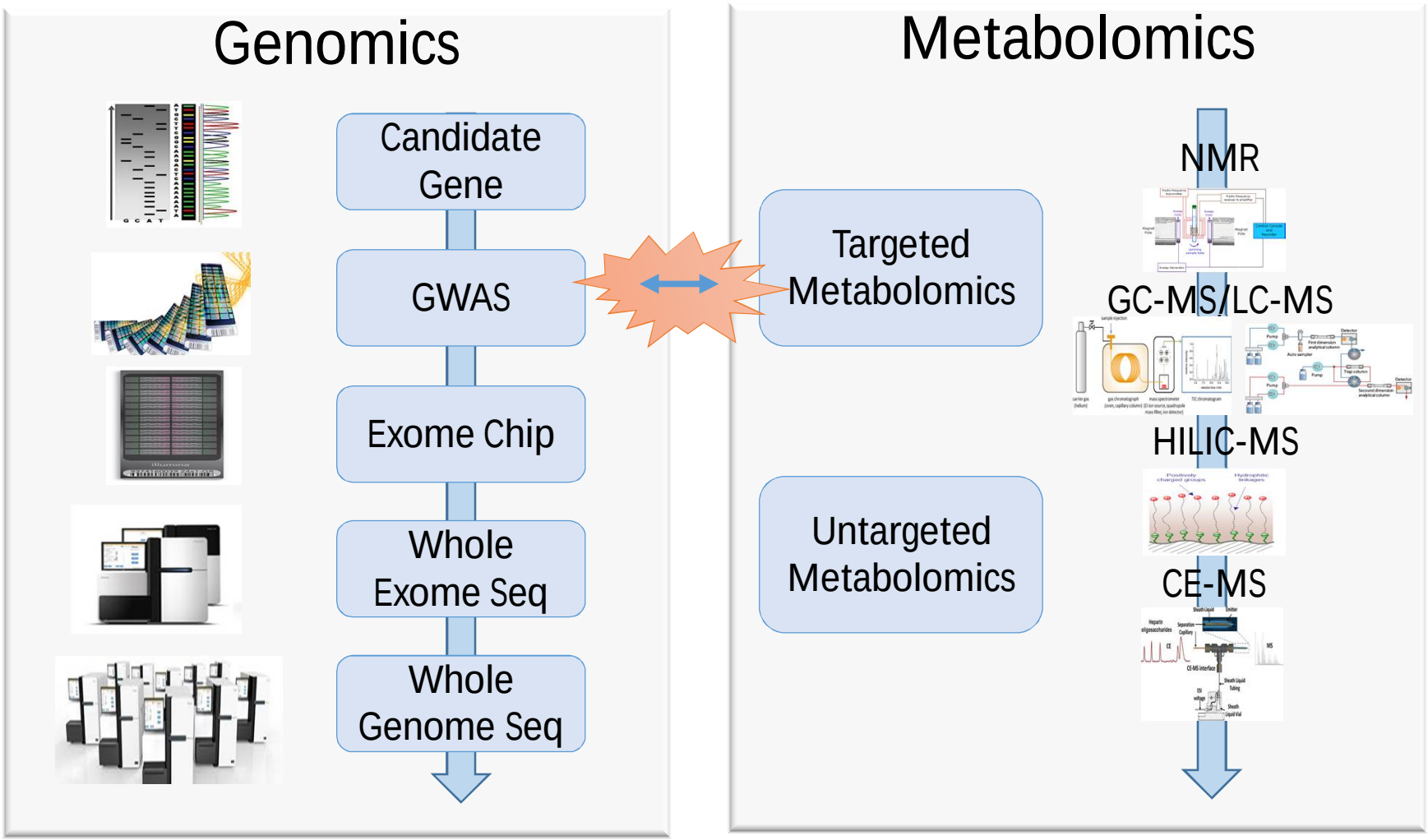
- Differences in Environment
- Differences in site frequency spectrum
- G x E
- Epidemiology of Disease





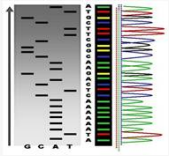


Advances in Genomics & Metabolome



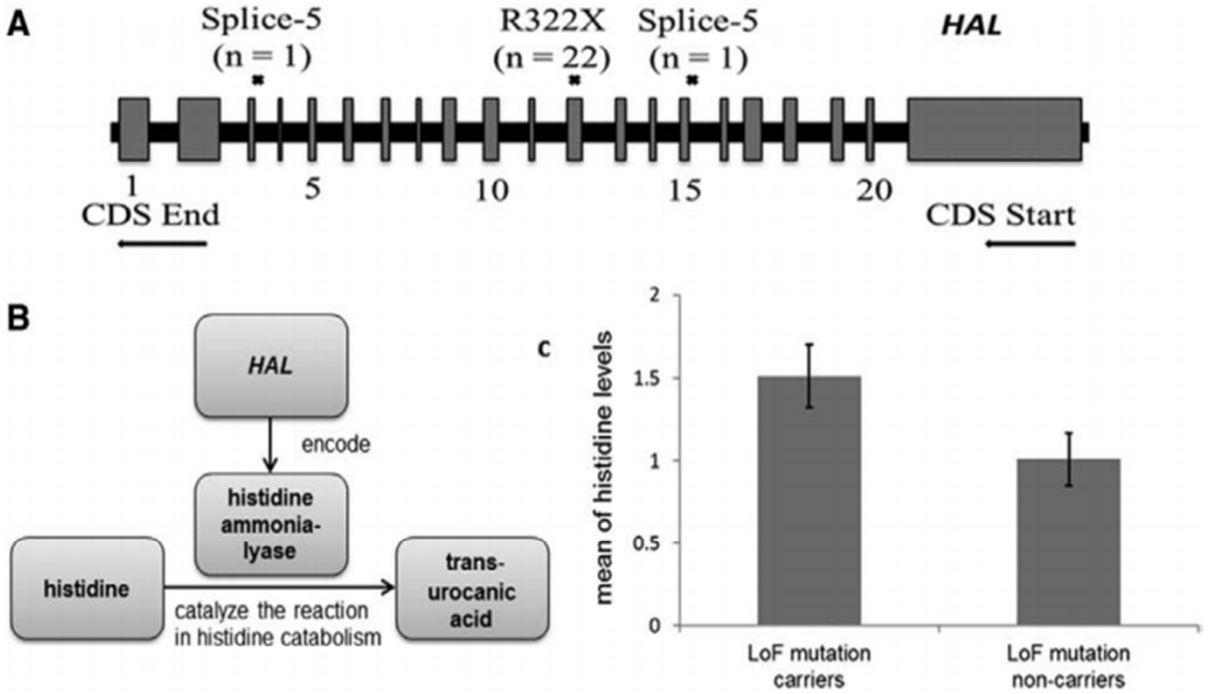
Advances in Genomics & Metabolome

Genomics



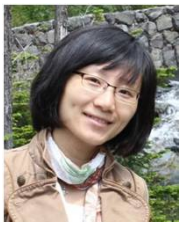
Candidate Gene

HAL, Histidine and Coronary Heart Disease



African Americans in ARIC

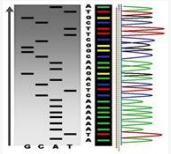
European Americans in FHS

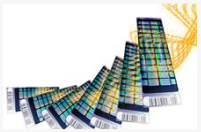


Yu, *Circ Cardiovasc Genet*, 2015

Advances in Genomics & Metabolome

Genomics





Candidate Gene

GWAS

GWAS on Metabolomics

A Genome-wide Association Study of the Human Metabolome in a Community-Based Cohort

Eugene P. Rhee,[1,6,18] Jennifer E. Ho,[8,11,18] Ming-Huei Chen,[9,1]
Martin G. Larson,[11,14] Anahita Ghorbani,[3] Xu Shi,[2] Iiro T. Helen
Amy Deik,[6] Kerry A. Pierce,[6] Kevin Bullock,[6] Geoffrey A. Walfo
Clary Clish,[6] J.-R. Joanna Yeh,[2] Thomas J. Wang,[16,17,19,*] and

ARTICLE

OPEN ACCESS Freely available online PLoS GENETICS

Genetics Meets Metabolomics: A Genome-Wide Association Study of Metabolite Profiles in Human Serum

Christian Gieger[1,2], Ludwig Geistlinger[1], Elisabeth Altmaier[3,4], Martin Hrabé de Angelis[5,6], Florian Kronenberg[7], Thomas Meitinger[8,9], Hans-Werner Mewes[3,10], H.-Erich Wichmann[1,2], Klaus M. Weinberger[11], Jerzy Adamski[5,6], Thomas Illig[1], Karsten Suhre[3,4]*

Human metabolic individuality in biomedical and pharmaceutical re

Karsten Suhre[1,2,3], So Youn Shin[4,*], Ann Kristin Petersen[5,*], Robert P. Mohney[6], David Meredith[7], Brigitu

OPEN ACCESS Freely available online PLOS GENETICS

Genetic Determinants Influencing Human Serum Metabolome among African Americans

Bing Yu[1], Yan Zheng[1], Danny Alexander[2], Alanna C. Morrison[1], Josef Coresh[3], Eric Boerwinkle[1,4]*

1 Human Genetics Center, University of Texas Health Science Center at Houston, Houston, Texas, United States of America, 2 Metabolon, Inc., Durham, North Carolina, United States of America, 3 Department of Epidemiology, Johns Hopkins University, Baltimore, Maryland, United States of America, 4 Human Genome Sequencing Center, Baylor College of Medicine, Houston, Texas, United States of America

A genome-wide association study in human urine

Karsten Suhre[1,2,18], Henri Wallaschofski[3], Johannes Raffler[1,2], Nele

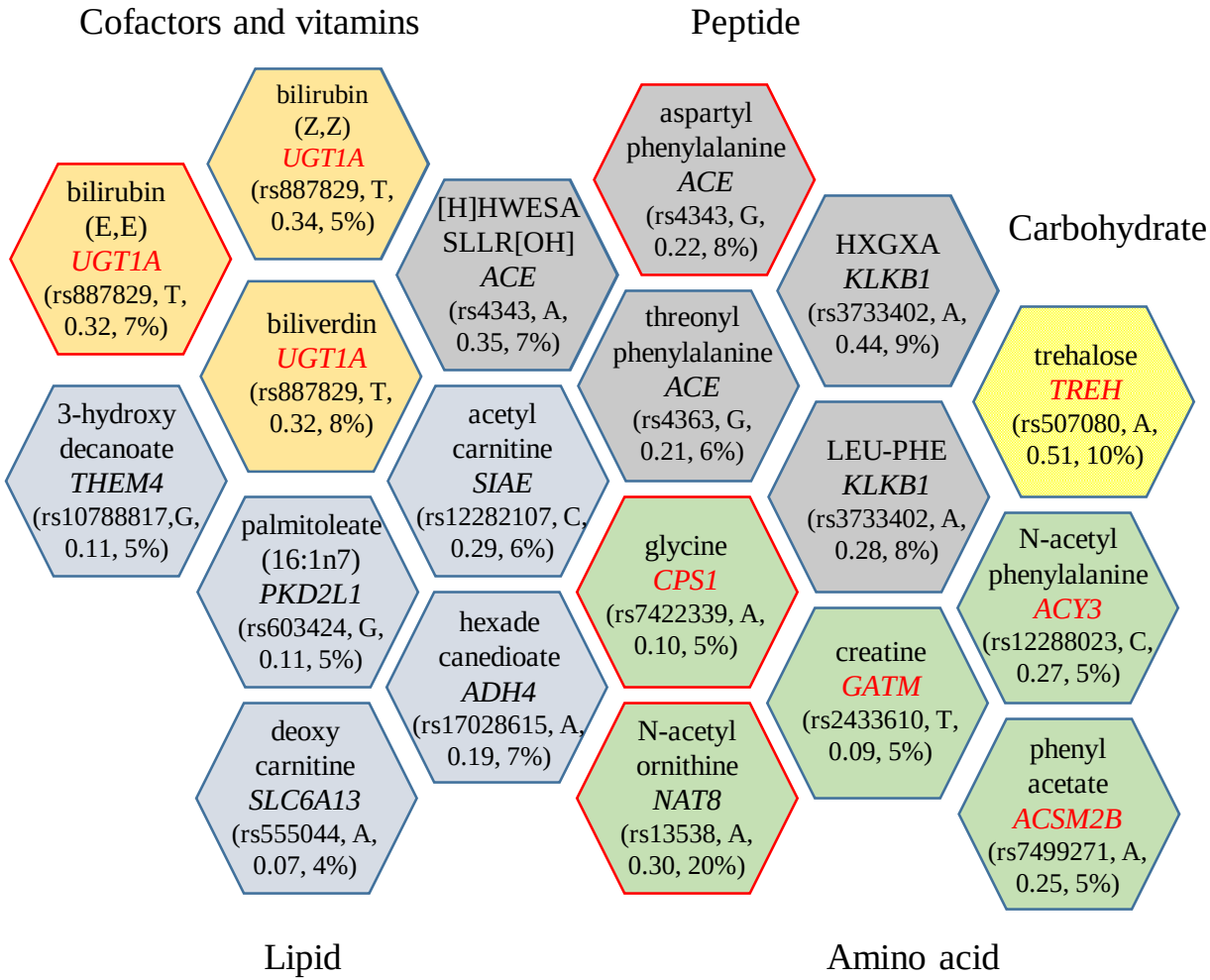
Genome-wide association study identifies multiple loci influencing human serum metabolite levels

Taru Tukiainen[1,3–5,37], Antti-Pekka Sarin[1,2], Alfredo Ortega-Alonso[1,6], Emmi Tikkanen[1,2], Antti J Kangas[5], Pasi Soininen[5,6], Peter Würtz[1,3,5], Kaisa Silander[1,2], Danielle M Dick[9], u J Savolainen[11,12], Jorma Viikari[13], Mika Kähönen[14], Terho Lehtimäki[7], Michael Inouye[17,18], Mark I McCarthy[19,20], Antti Jula[2], Johan Eriksson[21–24], ko Salomaa[2], Jaakko Kaprio[1,6,27], Marjo-Riitta Järvelin[3,12,28–30], Leena Peltonen[36], n B Freimer[32], Mika Ala-Korpela[5,6,11,12], Aarno Palotie[1,33–35] & Samuli Ripatti[1,2,33]

An atlas of genetic influences on human blood metabolites

So-Youn Shin[1,21,23], Eric B Fauman[2,23], Ann-Kristin Petersen[3,23], Jan Krumsiek[4,23], Rita Santos[5], Jie Huang[1], Matthias Arnold[6], Idil Erte[7], Vincenzo Forgetta[8], Tsun-Po Yang[1], Klaudia Walter[1], Cristina Menni[7], Lu Chen[1,9], Louella Vasquez[1], Ana M Valdes[7,10], Craig L Hyde[11], Vicky Wang[2], Daniel Ziemek[2], Phoebe Roberts[2,22], Li Xi[2], Elin Grundberg[6,12], The Multiple Tissue Human Expression Resource (MuTHER) Consortium[13], Melanie Waldenberger[14], J Brent Richards[7,8,15], Robert P Mohney[16], Michael V Milburn[16], Sally L John[17], Jeff Trimmer[18,21], Fabian J Theis[4,19], John P Overington[5], Karsten Suhre[6,20,24], M Julia Brosnan[11,24], Christian Gieger[3,24], Gabi Kastenmüller[6,24], Tim D Spector[7,24] & Nicole Soranzo[1,9,24]

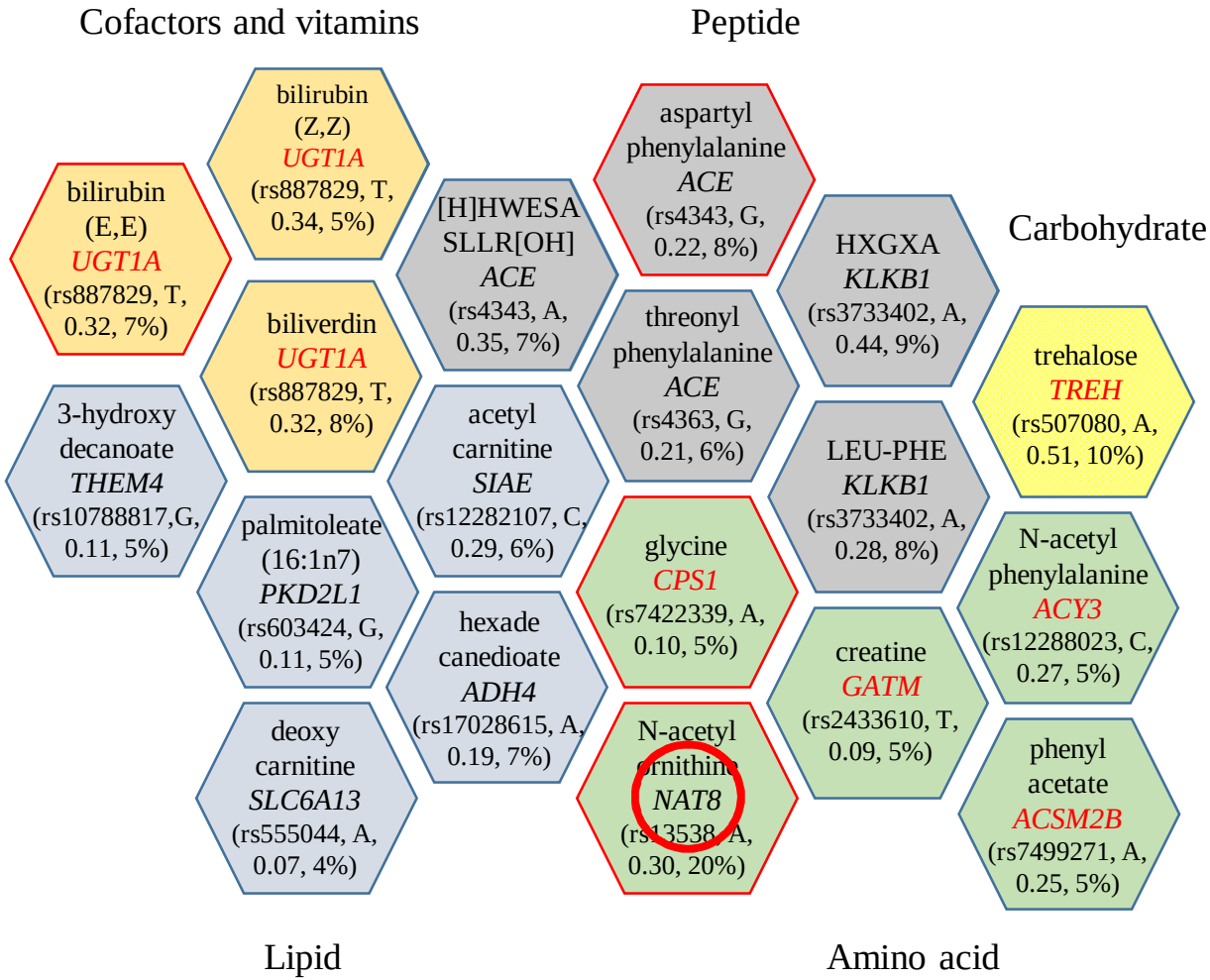
Genome-wide Significant Gene-Metabolite Pairs in 1,679 ARIC African Americans



Common variants with p-value < 1.6×10^{-10}

Yu, *PLoS Genet*, 2014

Genome-wide Significant Gene-Metabolite Pairs in 1,679 ARIC African Americans

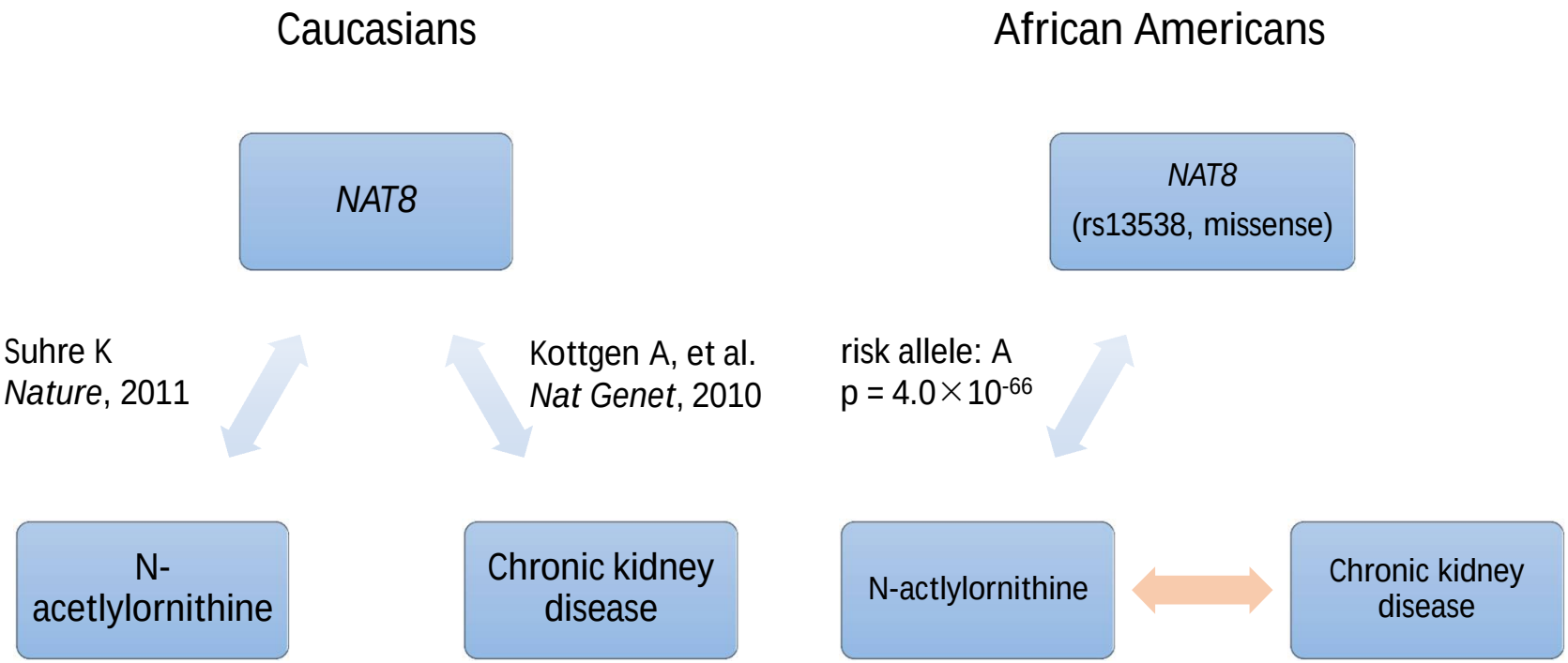


Common variants with p-value < 1.6×10^{-10}

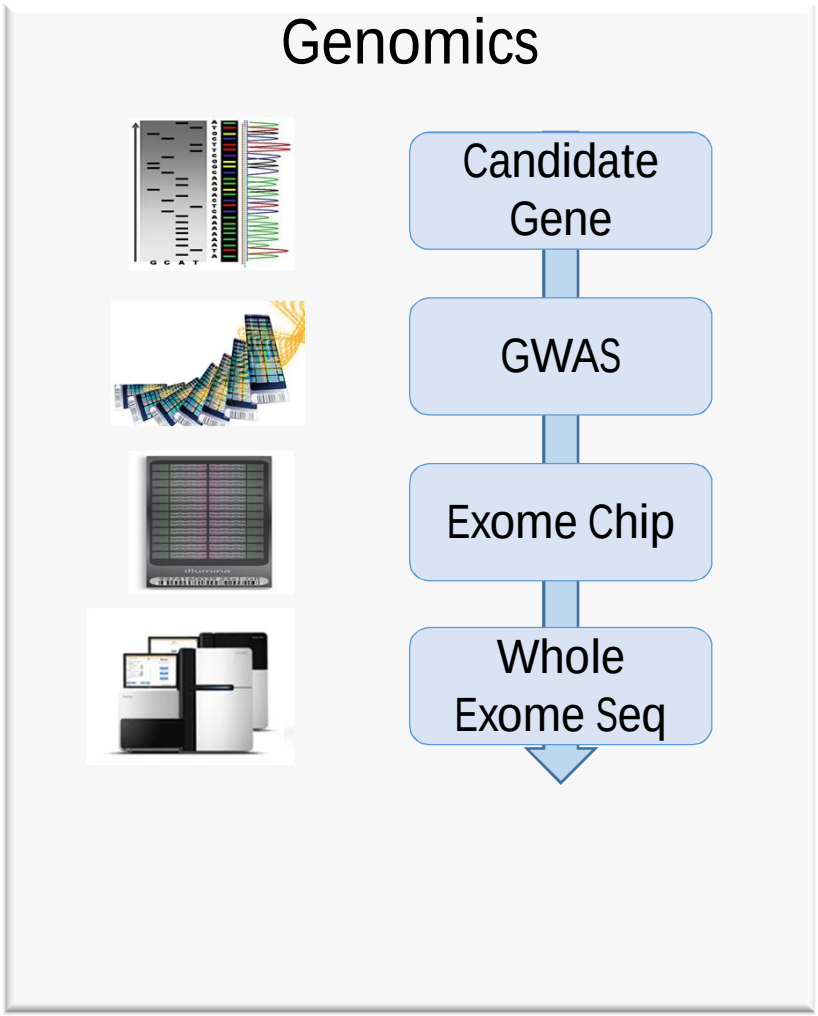
Yu, *PLoS Genet*, 2014

NAT8 and Chronic Kidney Disease

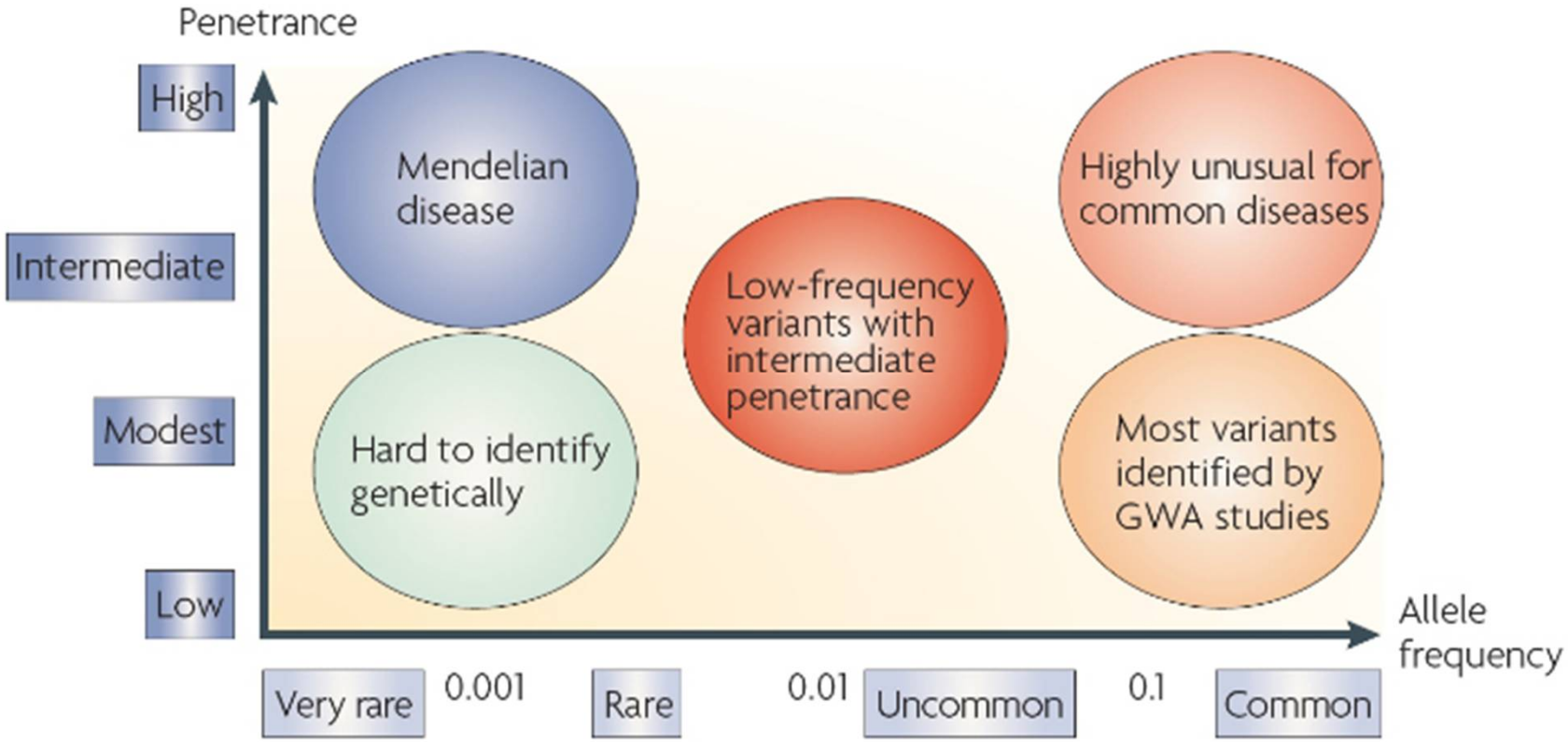
Hypothesis 1: *NAT8* – N-acetlyornithine – chronic kidney disease?



Advances in Genomics & Metabolome



Why Sequencing?



McCarthy, *Nat Reviews Genet*, 2009

Defining LOF

- Variants predicted to trigger nonsense-mediated decay (NMD)

- Categories:
  1) Premature stop codon-introducing
  2) Disrupt essential splice site
  3) Insertion/deletion frameshifts (indel)

- Additional Subdivision:
  - Full: all known protein coding transcripts
  - Partial: affecting only a fraction of known coding transcripts

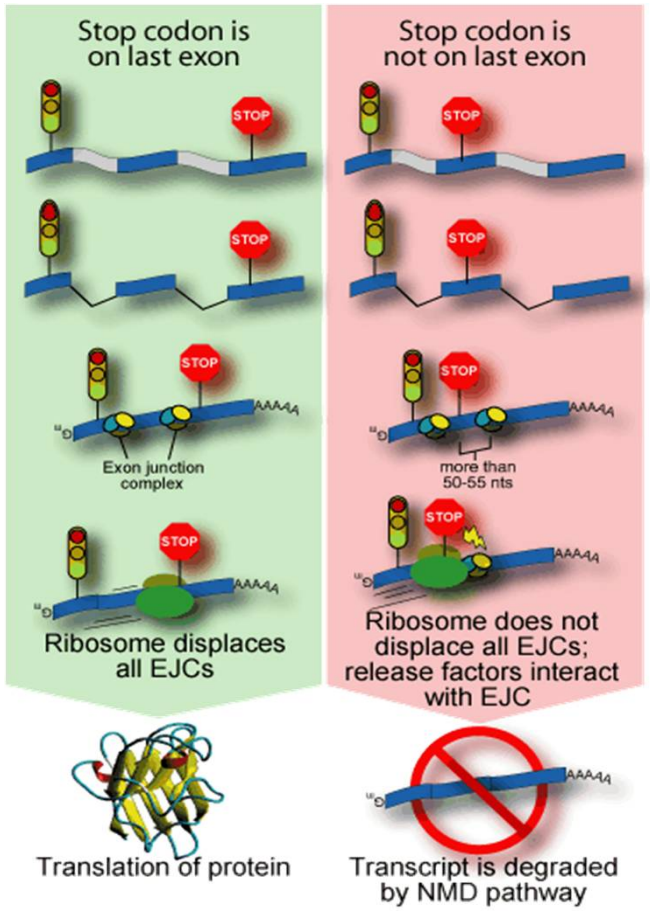


Image via: http://compbio.berkeley.edu/people/ed/rust/

Annotating LOF

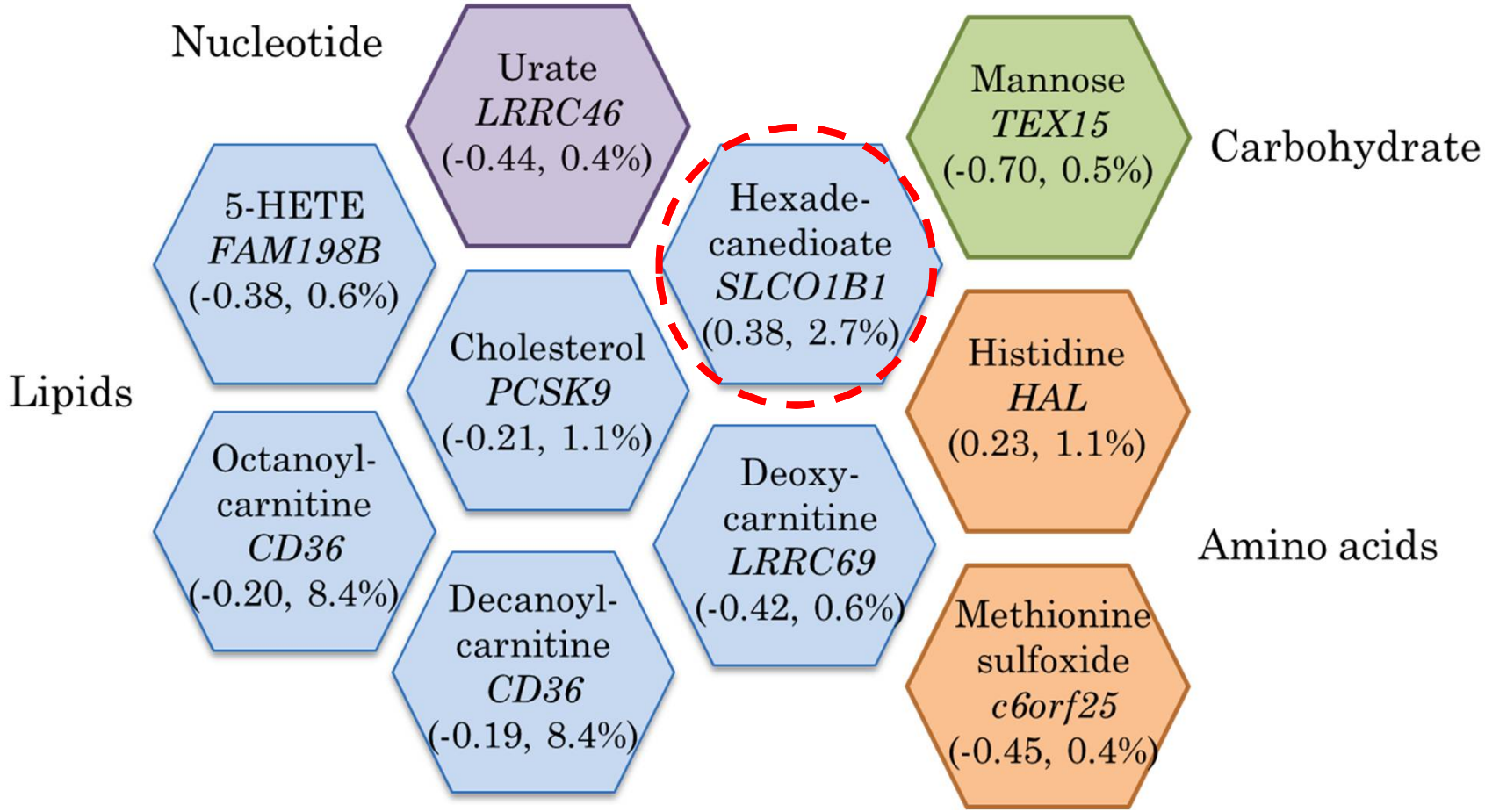
- 8,554 ARIC Study participants
  - 5,718 EA and 2,836 AA
  - (4,277 disc and 4,277 repl)
- Variant filtering:
  - Single-exon genes
  - Non protein-coding genes
  - Affect all gene isoforms
  - Terminal gene exon
- 36,787 LOF sites in 11,922 genes
- Average per individual:
  - Heterozygous (homozygous)

LOF type	Initial	After filtering	% Filtered out (Low-Confidence)
Stop gain	19,759	14,076	28.7%
Splice	10,634	8,843	16.8%
Frame Shift	33,703	13,868	58.8%
Total	**64,096**	**36,787**	

LOF type	AA	EA
Stop gain	27.3 (2.1)	21.1 (2.2)
Splice	16.7 (1.9)	9.6 (1.8)
Frame Shift	36.1 (4.4)	22.6 (3.1)
Total	**80.1 (8.4)**	**53.3 (7.1)**

FHS: phs000651.v4.p9; CHS: phs000667.v1.p1; ARIC phs000668.v1.p1

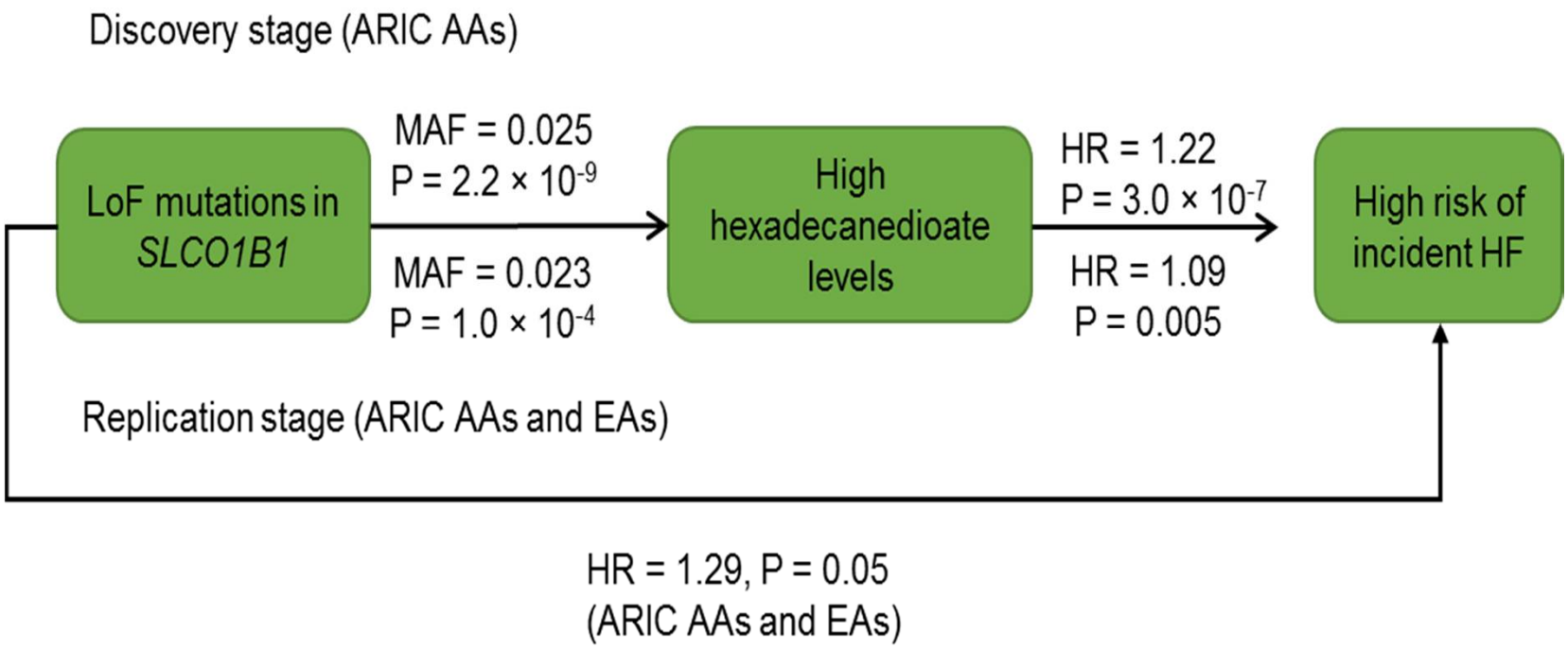
Significant LoF mutation metabolite pairs



324 single LoF varaints (MAF ≥ 5%),
1285 genes with cMAC ≥ 7 included,
p-value < 1.3×10^{-7}

Yu, *Sci Adv*, 2016

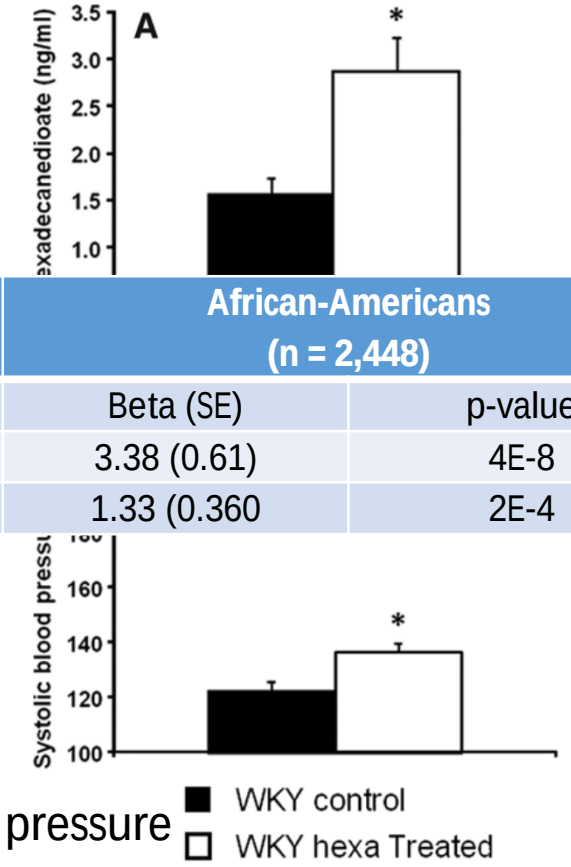
SLCO1B1, Hexadecanedioate & Heart Failure



Yu, *Sci Adv*, 2016

Possible Mechanism of the Association

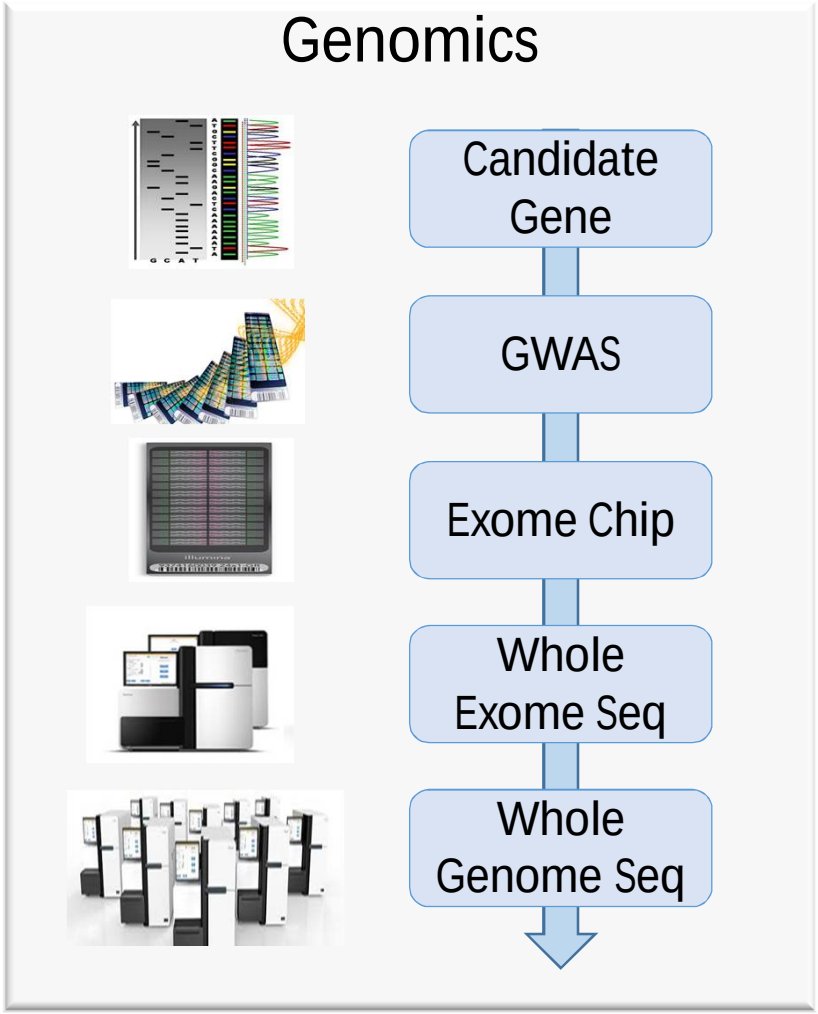
250 mg/kg/day hexadecanedioate feeding



	European-Americans (n = 1,551)		African-Americans (n = 2,448)	
	Beta (SE)	p-value	Beta (SE)	p-value
Systolic BP	1.30 (0.43)	0.002	3.38 (0.61)	4E-8
Diastolic BP	0.74 (0.26)	0.004	1.33 (0.360	2E-4

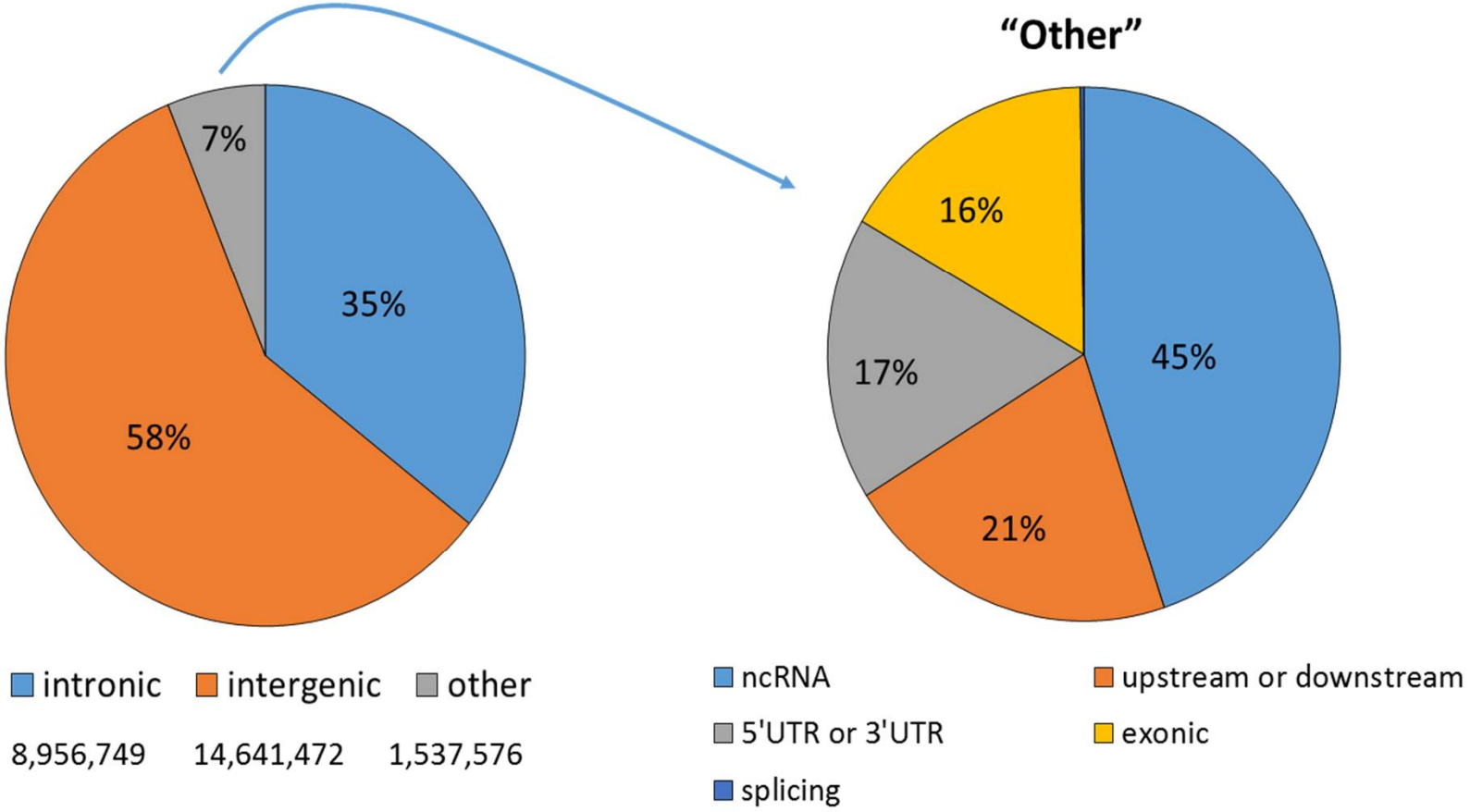
C. Menni *et al.*,
Metabolomic identification of a novel pathway of blood pressure regulation involving hexadecanedioate.
Hypertension **66**, 422 (Aug, 2015).

Advances in Genomics & Metabolome



Whole Genome Composition

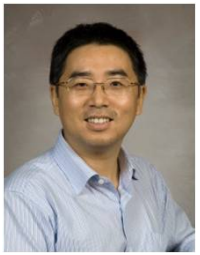
The genomes were annotated by ANNOVAR based on the RefSeq database



2.0×10^{15} sequenced bases

US corn production in 2014: 1.3×10^{15} kernels

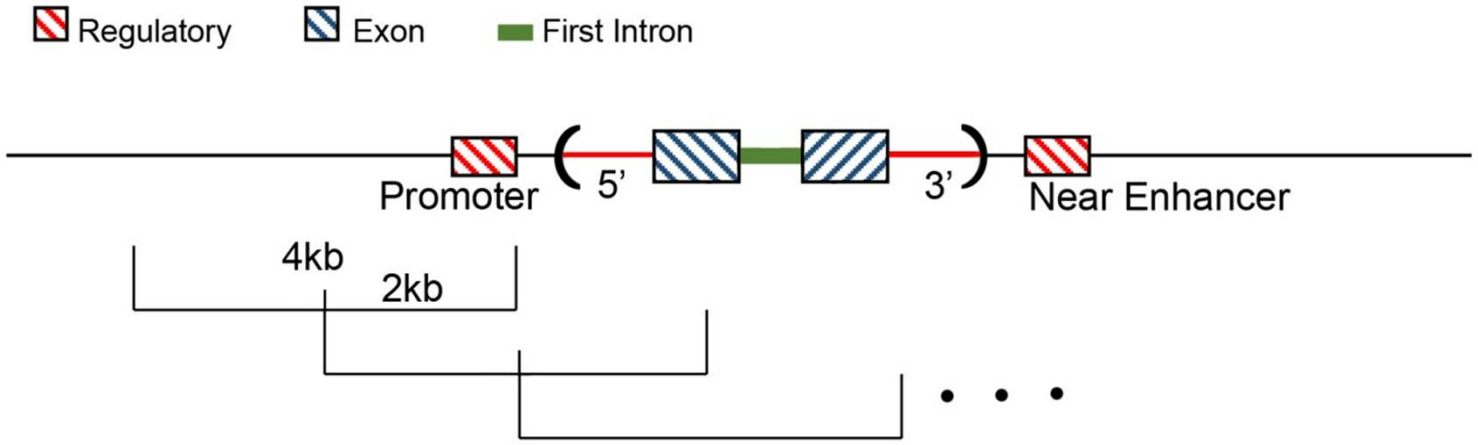
From G. Abecasis



Analytic Approach
WGS Annotation is Key

Aggregate by annotated functional motif for low frequency and rare SNVs (MAF ≤ 5%):

- Sliding window
- Regulatory domain
- First intron



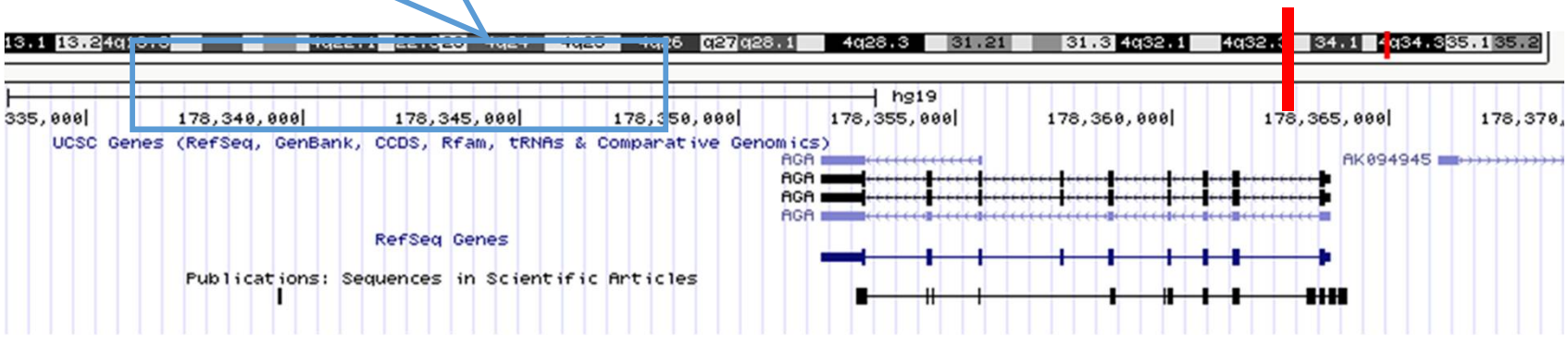
WGS Annotation Tool: WGSA (Liu et al, *J Med Genet*, 2015)

Asparagine:

- o A non-essential amino acid;
- o Biosynthesis/diet intake;
- o Required for development and function of the brain.

2 significant sliding windows (p < 4e-8), 6kb downstream of *AGA*

rs11131799, the most significant common variant



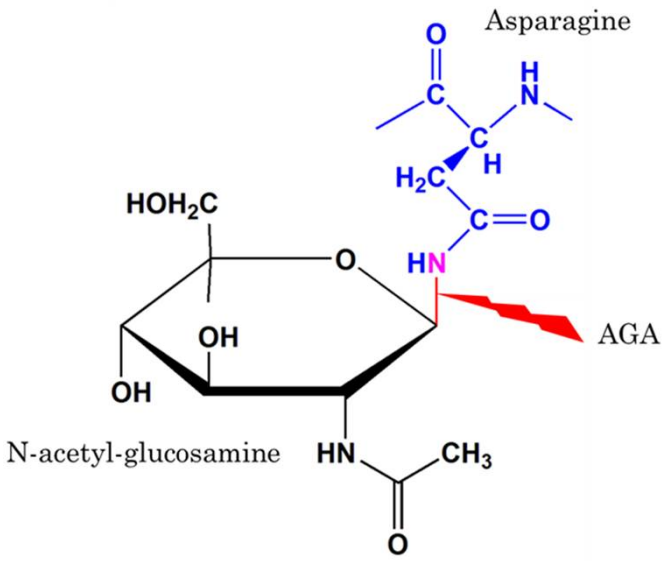
Functional coding variants (MAF < 5%) in *AGA* aggregately affect asparagine levels

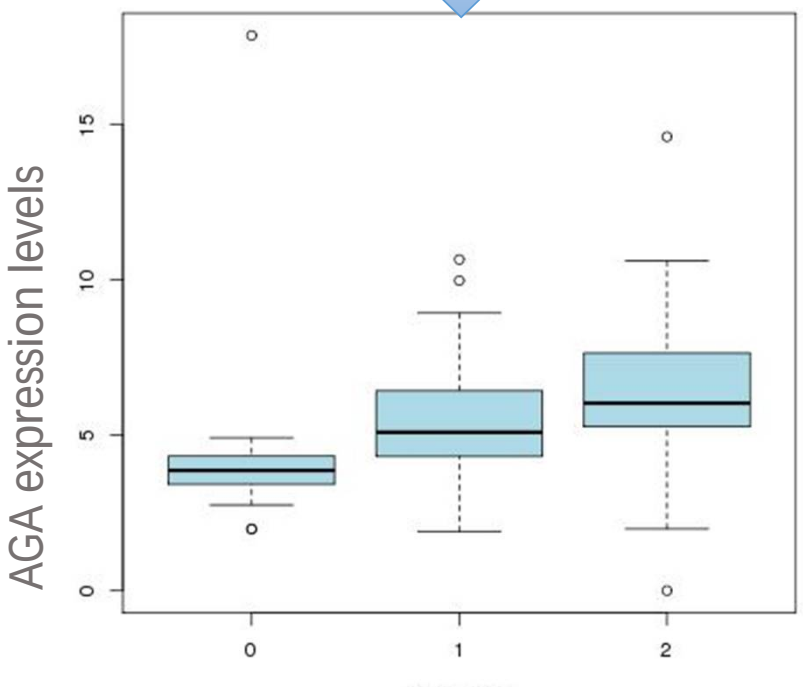
***AGA* gene:**

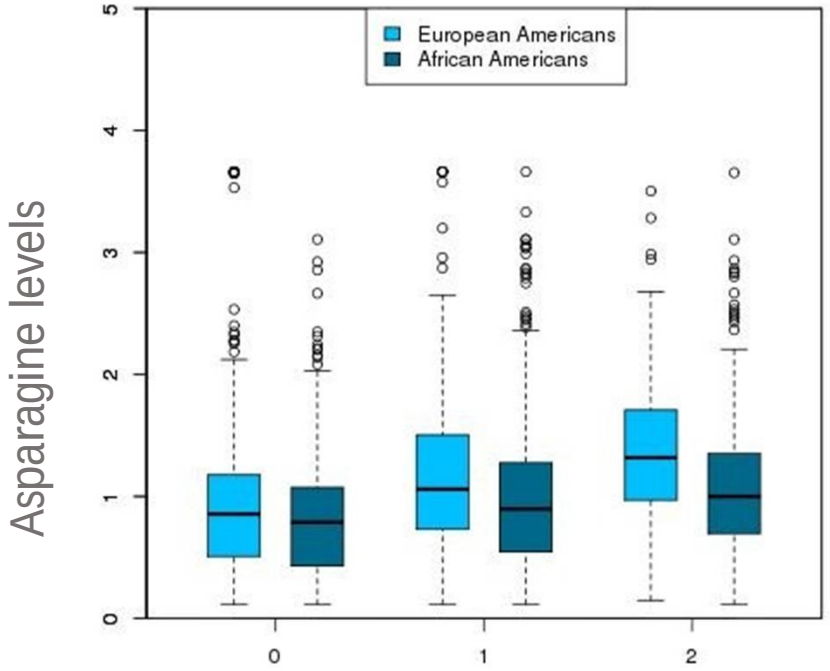
- Aspartylglucosaminidase
- cleaves *asparagine* from N-acetylglucosamine.

rs11131799 (1st intron of *AGA*):

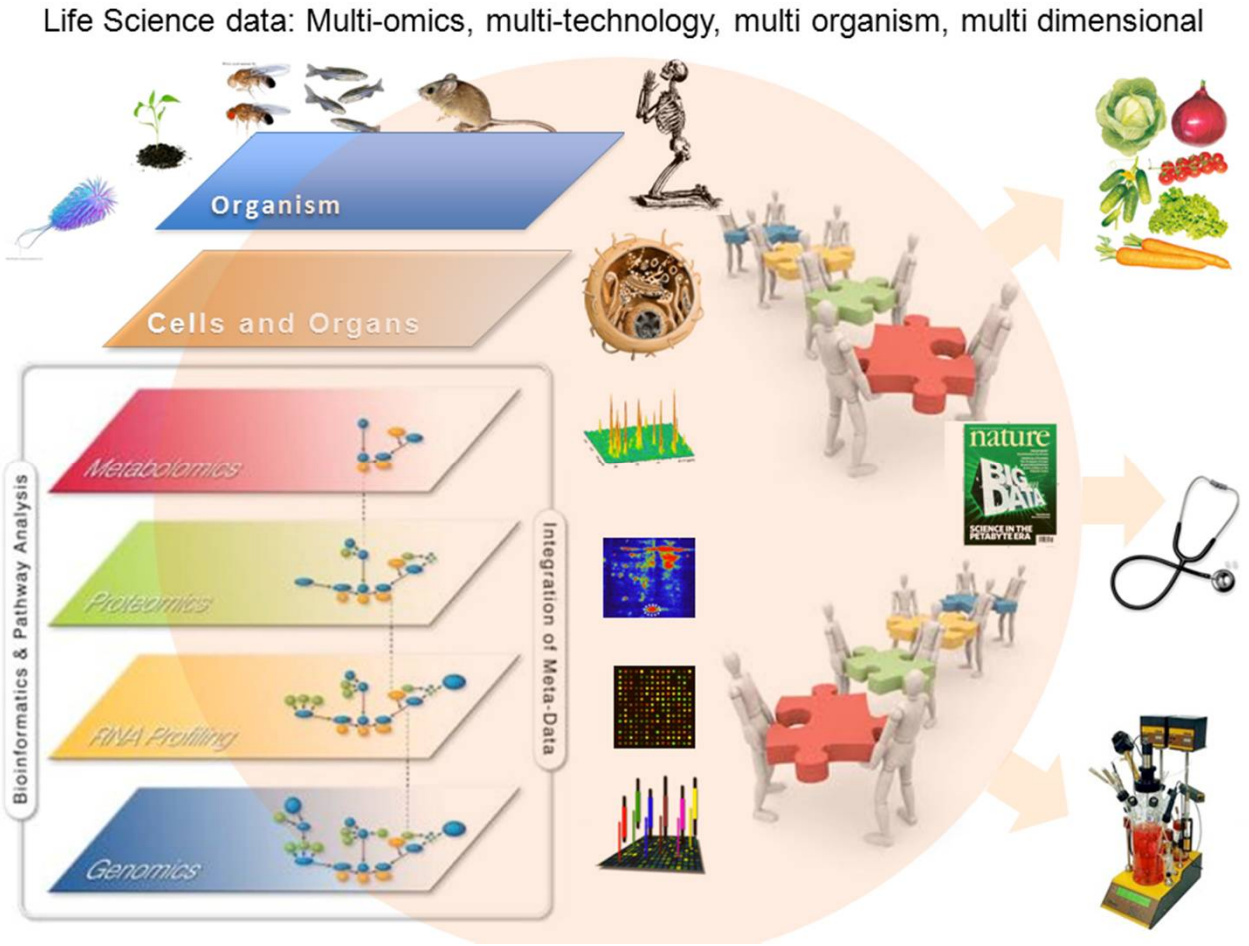
- Predicted promoter;
- Influences the expression levels of *AGA* ($p = 0.01$).







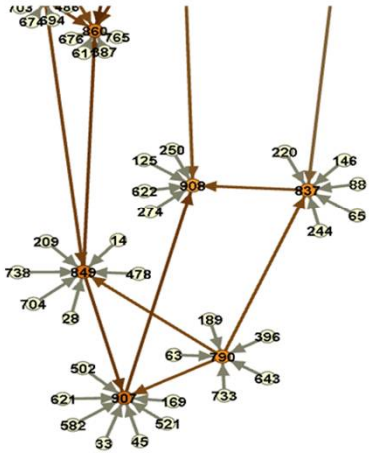
Multi-Omics Integration

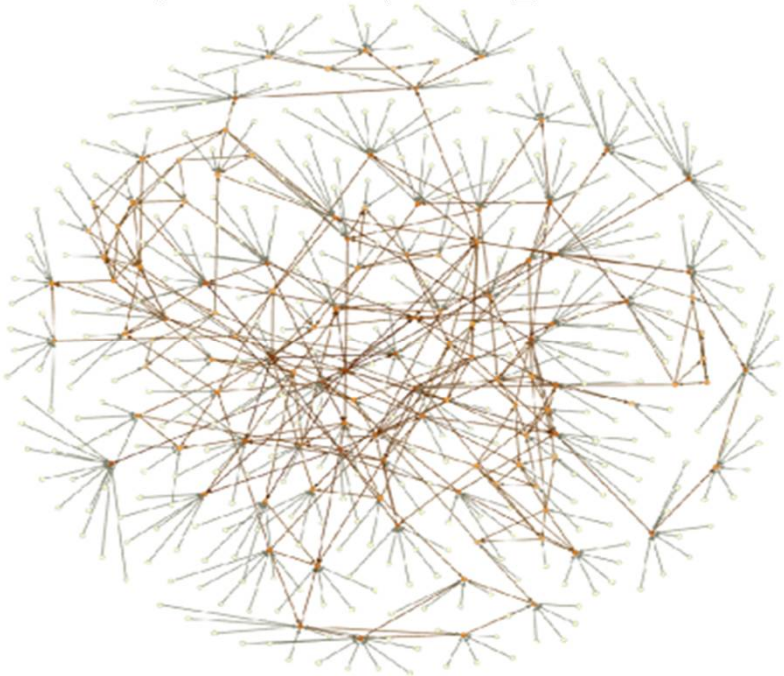


Achieving this vision, requires delivering large amounts of high quality data to the community in a timely manner.

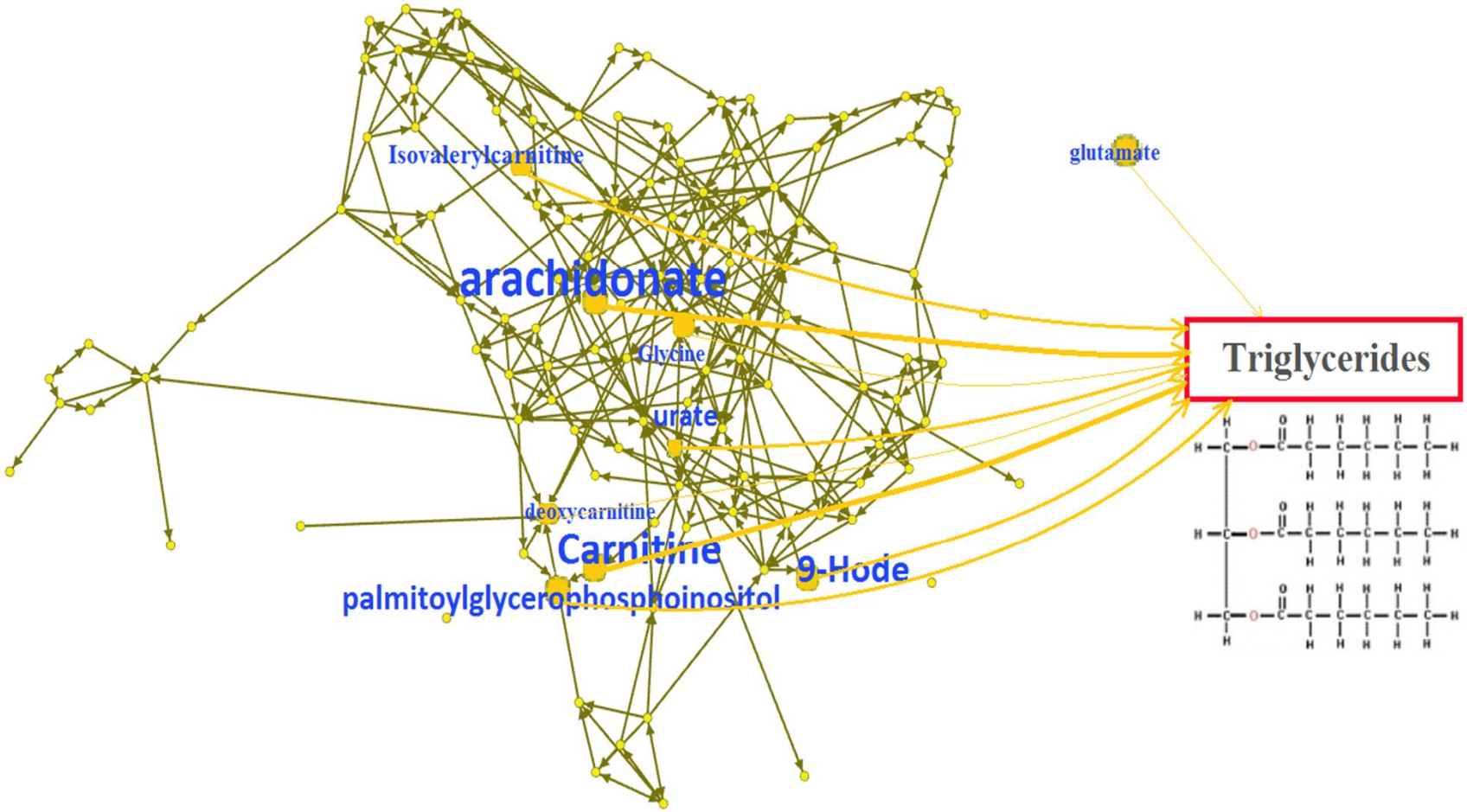
Identification of “Causal” Pathways among the Serum Metabolome

- As shown above, the principal of Mendelian randomization can lend credence to claims of causal inference.
- This principal of Mendelian randomization can extend to information across the genome.
- We (Yazdani, 2016) have combined the principal of genome-wide Mendelian randomization with Directed Acyclic Graph algorithms. (GDAG).





Metabolomic-Triglycerides Network



In total, 9 metabolites have direct a effect on triglyceride levels.

Yazdani et. al, (2016). Metabolomics

Acknowledgments

University of Texas Health Science Center at Houston
Bing Yu
Mandana Yazdani
Elena Feofanova
Paul de Vries
Xiaoming Liu
Alanna Morrison

Baylor College of Medicine
Ginger Metcalf
Fuli Yu
Donna Muzny
Richard Gibbs

Framingham Heart Study
Robert Gerszten
Eugene Rhee
Dan Levy
Vasan Ramachandran

Atherosclerosis Risk in Communities Study
Christie Ballantyne
Josef Coresh

METABOLON
SOLUTIONS BEYOND THE DATA...™