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[8,12], 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]

NATURE GENETICS VOLUME 46 | NUMBER 6 | JUNE 2014

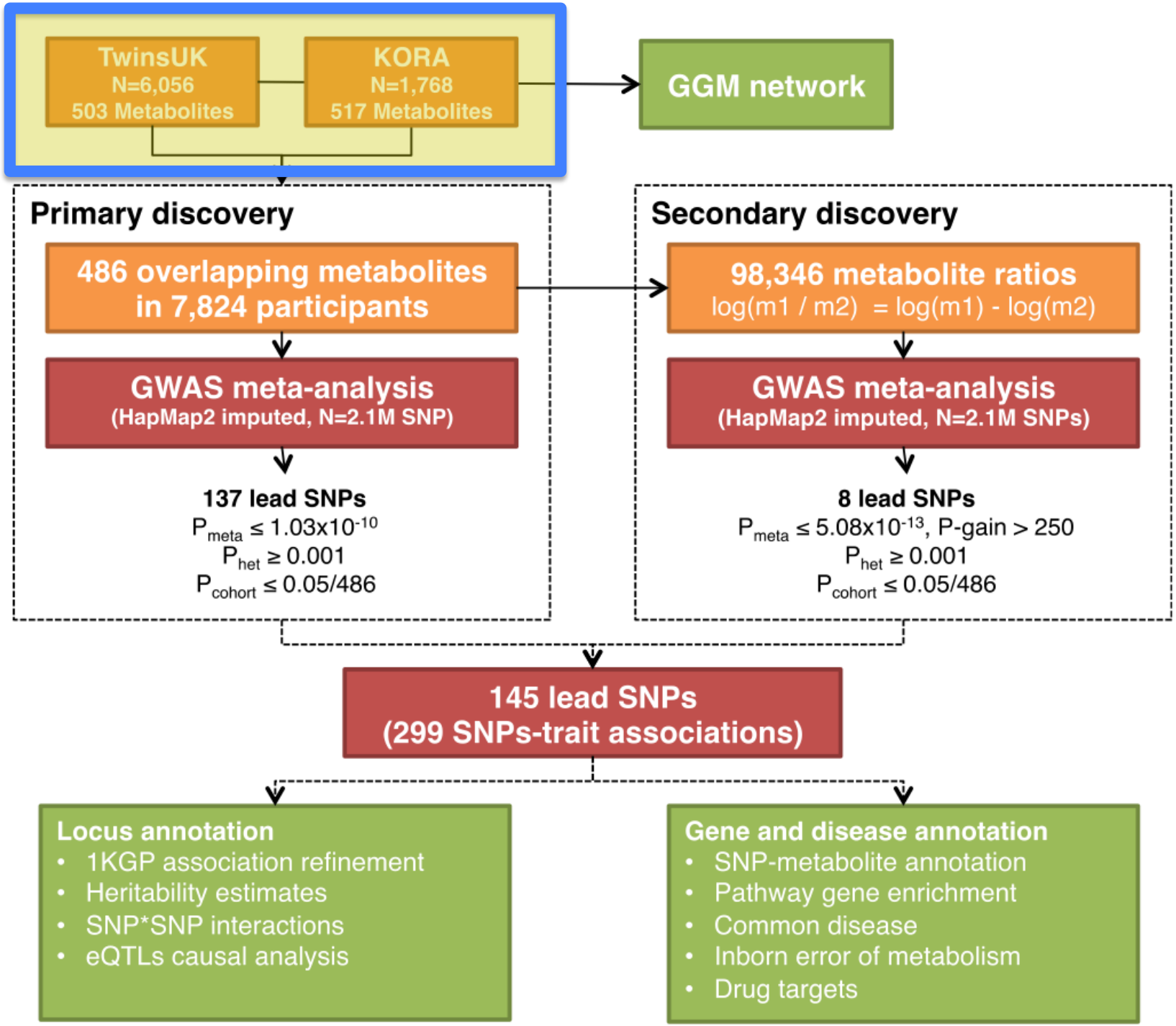
Presented by: Satrio Husodo
9/2014

Major goal of the study

- To identify genetic factors that control metabolism
  - Inborn errors of metabolism (PKU, MCAD deficiency, etc.)
  - Cardiovascular disease
  - Energy conversion
- Use metabolites as biomarkers
  - Use to assess drug response
- Atlas of gene-metabolite network can direct strategies in correcting deregulation of metabolism

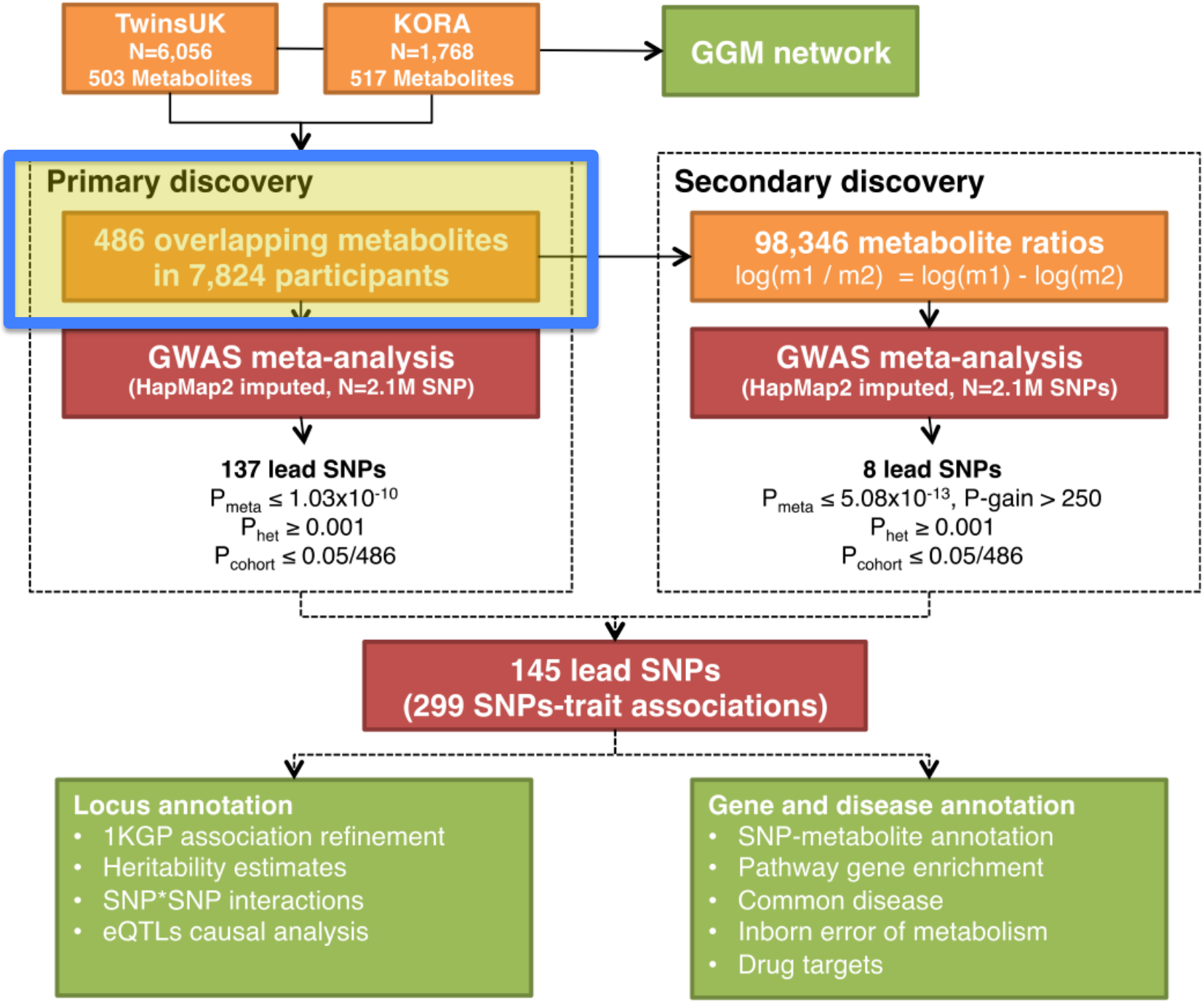
Experimental design overview

- Surveyed genome loci that are associated with metabolic traits
  - Genome-wide association studies (GWAS)
  - Metabolic profiling
- Atlas of blood metabolome
  - Network of genes-metabolites
- Quantified heritability, epistatic interactions of the genetic factors.


TwinsUK
N=6,056
503 Metabolites
KORA
N=1,768
517 Metabolites
GGM network
Primary discovery
486 overlapping metabolites in 7,824 participants
GWAS meta-analysis
(HapMap2 imputed, N=2.1M SNP)
137 lead SNPs
$P_{meta} \leq 1.03x10^{-10}$
$P_{het} \geq 0.001$
$P_{cohort} \leq 0.05/486$
Secondary discovery
98,346 metabolite ratios
log(m1 / m2) = log(m1) - log(m2)
GWAS meta-analysis
(HapMap2 imputed, N=2.1M SNPs)
8 lead SNPs
$P_{meta} \leq 5.08x10^{-13}$, P-gain > 250
$P_{het} \geq 0.001$
$P_{cohort} \leq 0.05/486$
145 lead SNPs
(299 SNPs-trait associations)
Locus annotation
• 1KGP association refinement
• Heritability estimates
• SNP*SNP interactions
• eQTLs causal analysis
Gene and disease annotation
• SNP-metabolite annotation
• Pathway gene enrichment
• Common disease
• Inborn error of metabolism
• Drug targets

Study samples

- KORA (Cooperative Health Research in the Region of Augsburg)
  - A series of independent epidemiological surveys in Germany
  - Mean age of 61 years
- TwinsUK
  - Adult UK twin registry
  - 93% women, mean age of 51 years


TwinsUK
N=6,056
503 Metabolites
KORA
N=1,768
517 Metabolites
GGM network
Primary discovery
486 overlapping metabolites
in 7,824 participants
GWAS meta-analysis
(HapMap2 imputed, N=2.1M SNP)
137 lead SNPs
$P_{meta} \leq 1.03x10^{-10}$
$P_{het} \geq 0.001$
$P_{cohort} \leq 0.05/486$
Secondary discovery
98,346 metabolite ratios
log(m1 / m2) = log(m1) - log(m2)
GWAS meta-analysis
(HapMap2 imputed, N=2.1M SNPs)
8 lead SNPs
$P_{meta} \leq 5.08x10^{-13}$, P-gain > 250
$P_{het} \geq 0.001$
$P_{cohort} \leq 0.05/486$
145 lead SNPs
(299 SNPs-trait associations)
Locus annotation
• 1KGP association refinement
• Heritability estimates
• SNP*SNP interactions
• eQTLs causal analysis
Gene and disease annotation
• SNP-metabolite annotation
• Pathway gene enrichment
• Common disease
• Inborn error of metabolism
• Drug targets

Metabolic data collection

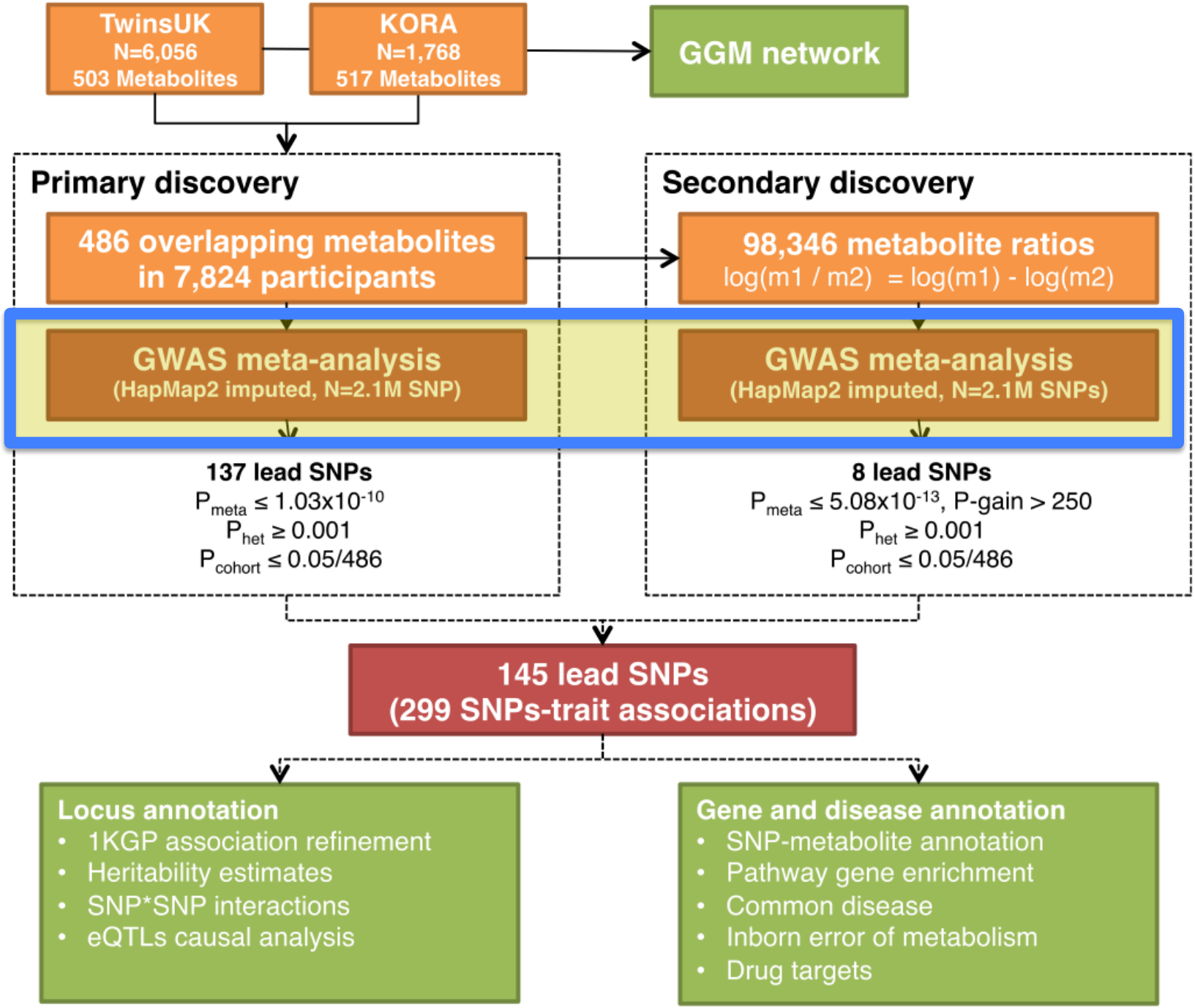
- Sampling
  - Plasma and serum collected post-examination
- Mass Spectrometry
  - UPLC-MS: Ultra High Performance Liquid Chromatography-Tandem Mass Spectrometry
  - GC-MS: Gas Chromatography-Mass Spectrometry
- Controls: Pooled human plasma and external standards

Compound identification and quantification

- Reference library of 4,000 chemically known standards
- 5,300 additional spectral entries of unknown structures
- Peaks are quantified using area under the curve

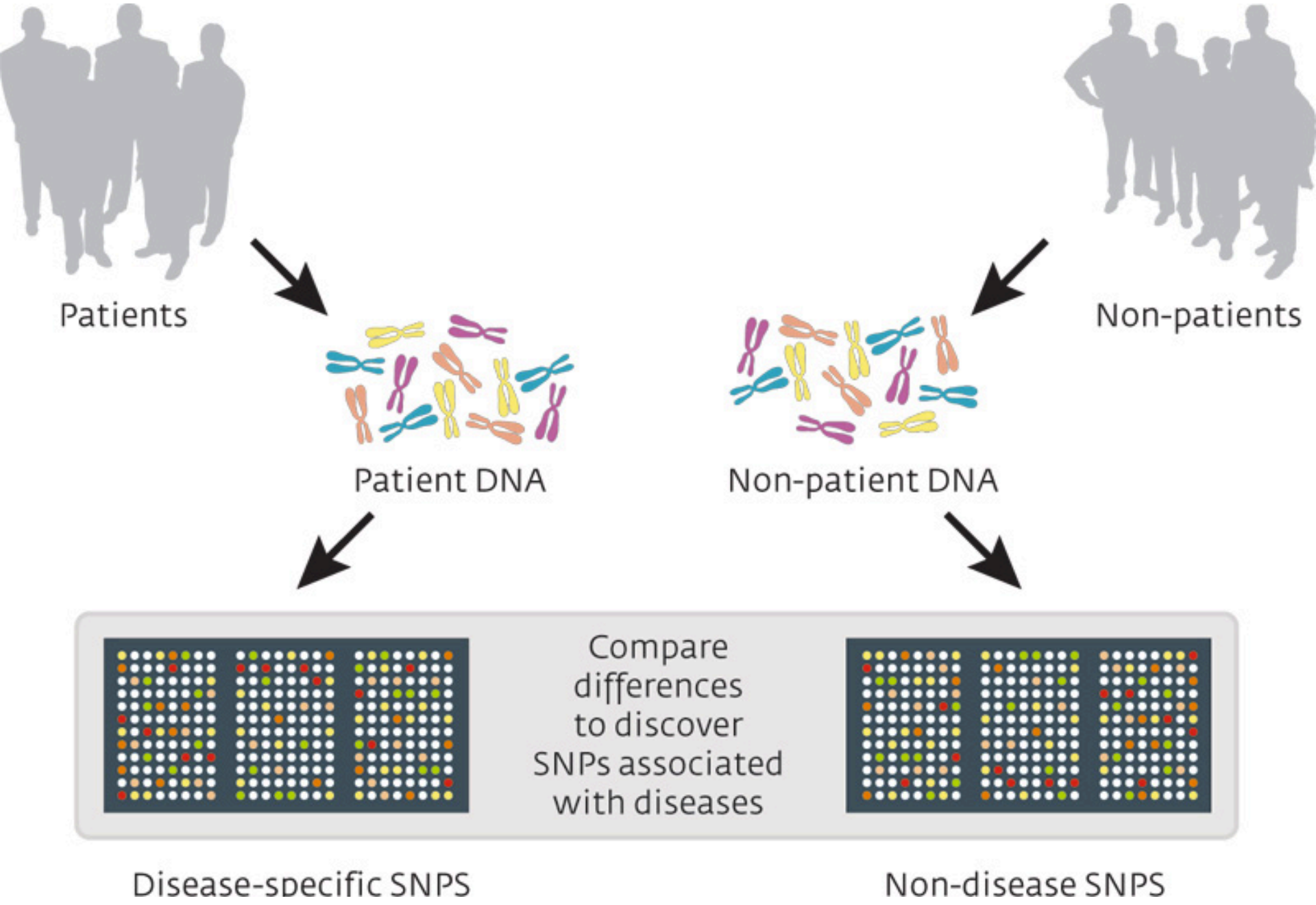
Metabolic profiling results

- 529 metabolites match library compounds
  - 63% of known structure
- 486 metabolites pass quality control and are monitored in subsequent genetic analysis
- Divided into 8 metabolic groups as annotated by KEGG database
  - 63 distinct biochemical pathways


TwinsUK
N=6,056
503 Metabolites
KORA
N=1,768
517 Metabolites
GGM network
Primary discovery
Secondary discovery
486 overlapping metabolites in 7,824 participants
98,346 metabolite ratios
log(m1 / m2) = log(m1) - log(m2)
GWAS meta-analysis
(HapMap2 imputed, N=2.1M SNP)
GWAS meta-analysis
(HapMap2 imputed, N=2.1M SNPs)
137 lead SNPs
$P_{meta} \leq 1.03x10^{-10}$
$P_{het} \geq 0.001$
$P_{cohort} \leq 0.05/486$
8 lead SNPs
$P_{meta} \leq 5.08x10^{-13}$, P-gain > 250
$P_{het} \geq 0.001$
$P_{cohort} \leq 0.05/486$
145 lead SNPs
(299 SNPs-trait associations)
Locus annotation
• 1KGP association refinement
• Heritability estimates
• SNP*SNP interactions
• eQTLs causal analysis
Gene and disease annotation
• SNP-metabolite annotation
• Pathway gene enrichment
• Common disease
• Inborn error of metabolism
• Drug targets

GWAS overview

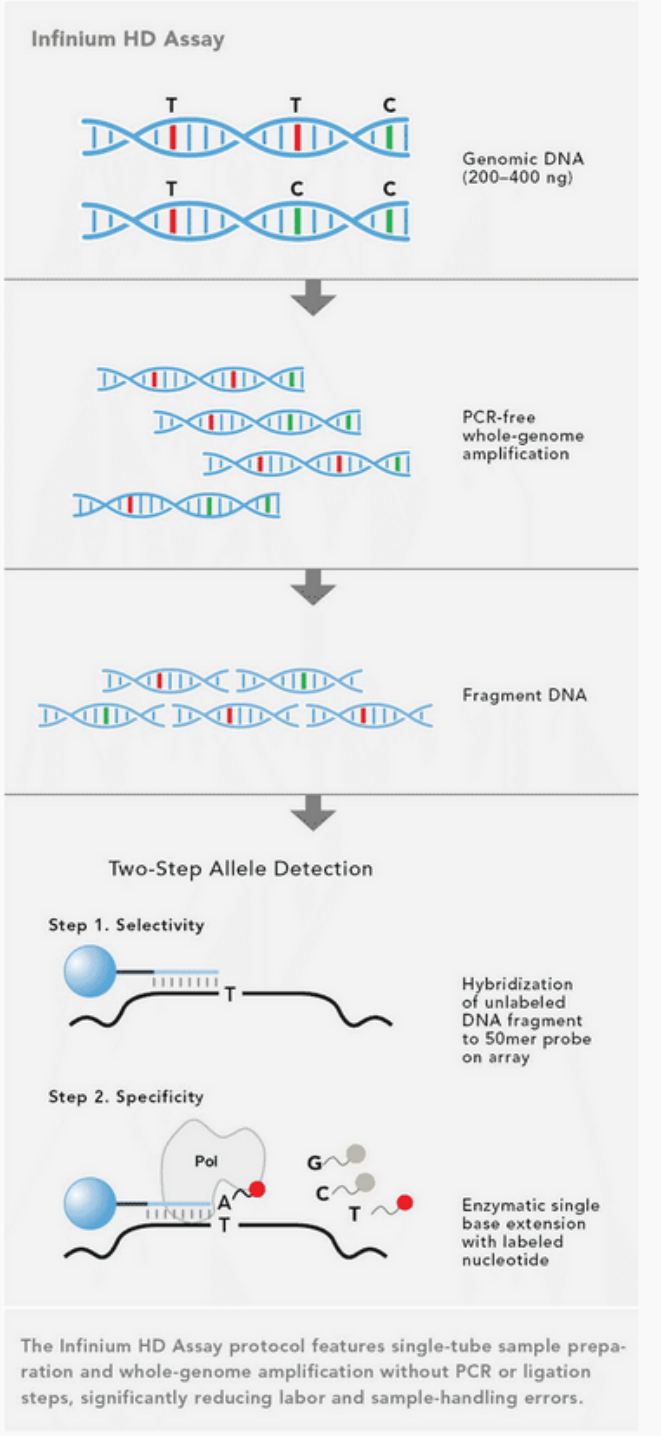
- Genome-wide association study
  - Goal: To identify genetic causes of a phenotype or disease
  - Analysis of genetic markers present in a population
  - Identifies association between markers/gene loci/genes to a particular trait
- *This study associates **single nucleotide polymorphism (SNP)** with **metabolite level***



http://www.mpg.de/10680/Modern_psychiatry

SNP array

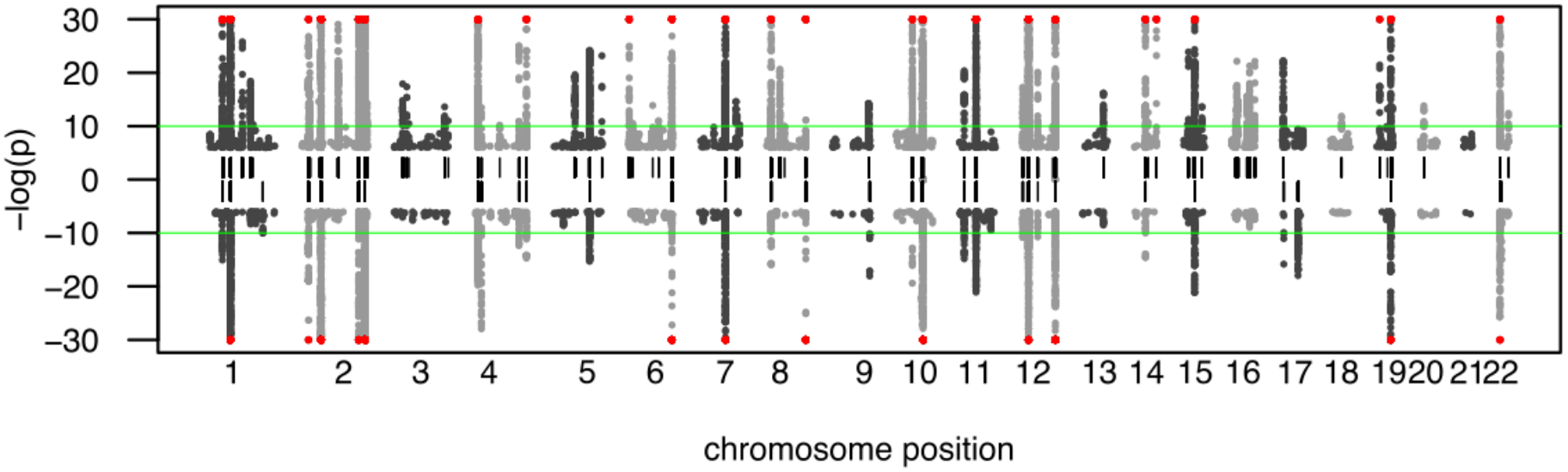
- SNPs identified using Illumina HD assay
- Genomic DNA amplified, fragmented
- Array contains beads that have probes to specific loci
- Single base extension with labeled nucleotide



Loci identification

- Genotyped for 2.1 million SNPs
- Initial discovery: 137 SNPs found to associate with metabolite concentration
- Analyzed 98,000 pairwise metabolite ratios
  - Found 8 additional loci associated with ratios
- Total: 299 SNP-metabolite associations
  - 145 statistically independent SNPs
  - 84 novel findings

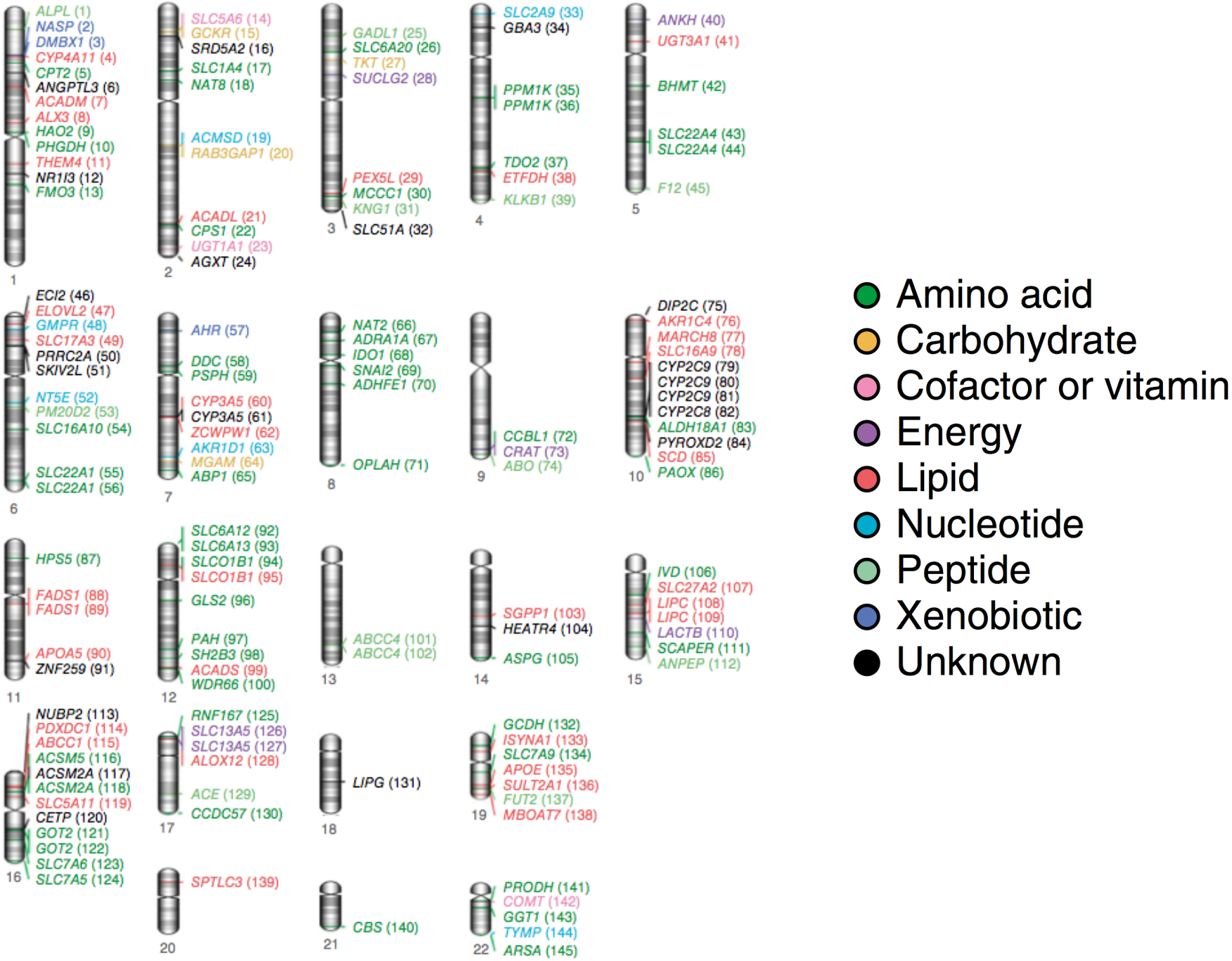
Association results

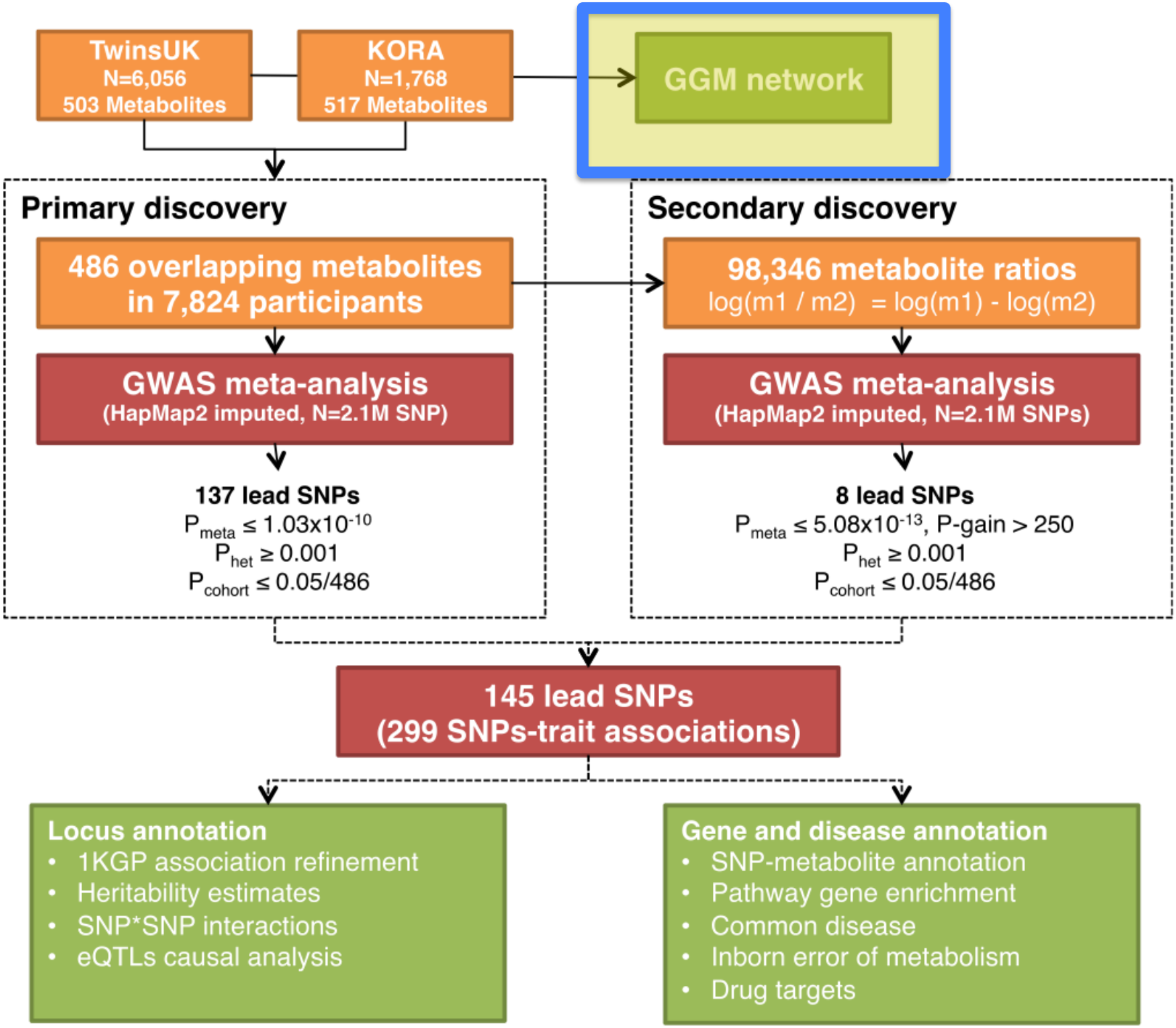


Top: TwinsUK

Bottom: KORA

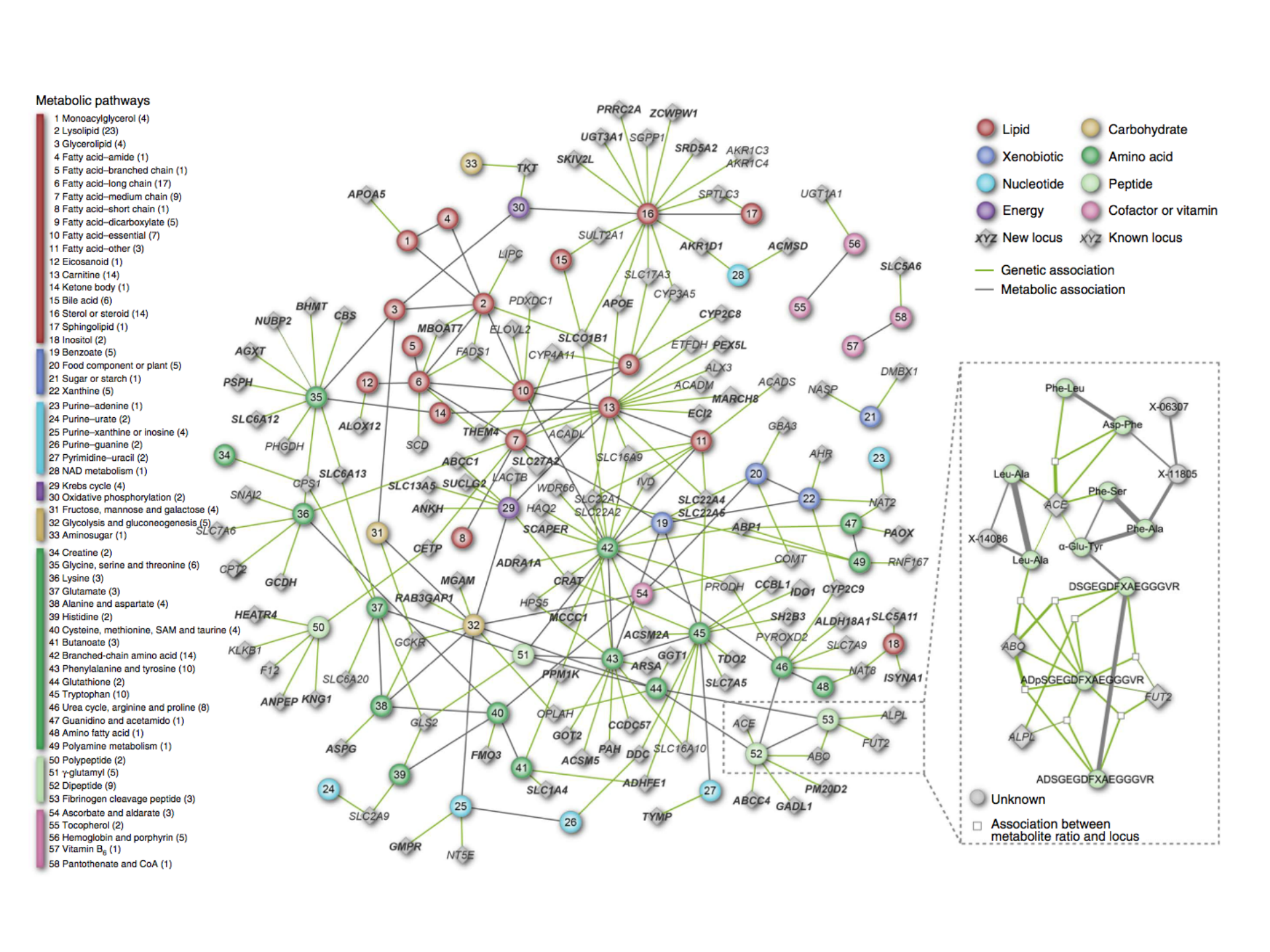
Green line: P-value cutoff (1 X 10^{-10})


ALPL (1)
NASP (2)
DMBX1 (3)
CYP4A11 (4)
CPT2 (5)
ANGPTL3 (6)
ACADM (7)
ALX3 (8)
HAO2 (9)
PHGDH (10)
THEM4 (11)
NR1I3 (12)
FMO3 (13)
1
SLC5A6 (14)
GCKR (15)
SRD5A2 (16)
SLC1A4 (17)
NAT8 (18)
ACMSD (19)
RAB3GAP1 (20)
ACADL (21)
CPS1 (22)
UGT1A1 (23)
AGXT (24)
2
GADL1 (25)
SLC6A20 (26)
TKT (27)
SUCLG2 (28)
PEX5L (29)
MCCC1 (30)
KNG1 (31)
SLC51A (32)
3
SLC2A9 (33)
GBA3 (34)
PPM1K (35)
PPM1K (36)
TDO2 (37)
ETFDH (38)
KLKB1 (39)
4
ANKH (40)
UGT3A1 (41)
BHMT (42)
SLC22A4 (43)
SLC22A4 (44)
F12 (45)
5
ECI2 (46)
ELOVL2 (47)
GMPR (48)
SLC17A3 (49)
PRRC2A (50)
SKIV2L (51)
NT5E (52)
PM20D2 (53)
SLC16A10 (54)
SLC22A1 (55)
SLC22A1 (56)
6
AHR (57)
DDC (58)
PSPH (59)
CYP3A5 (60)
CYP3A5 (61)
ZCWPW1 (62)
AKR1D1 (63)
MGAM (64)
ABP1 (65)
7
NAT2 (66)
ADRA1A (67)
IDO1 (68)
SNAI2 (69)
ADHFE1 (70)
OPLAH (71)
8
CCBL1 (72)
CRAT (73)
ABO (74)
9
DIP2C (75)
AKR1C4 (76)
MARCH8 (77)
SLC16A9 (78)
CYP2C9 (79)
CYP2C9 (80)
CYP2C9 (81)
CYP2C8 (82)
ALDH18A1 (83)
PYROXD2 (84)
SCD (85)
PAOX (86)
10
HPS5 (87)
FADS1 (88)
FADS1 (89)
APOA5 (90)
ZNF259 (91)
11
SLC6A12 (92)
SLC6A13 (93)
SLCO1B1 (94)
SLCO1B1 (95)
GLS2 (96)
PAH (97)
SH2B3 (98)
ACADS (99)
WDR66 (100)
12
ABCC4 (101)
ABCC4 (102)
13
SGPP1 (103)
HEATR4 (104)
ASPG (105)
14
IVD (106)
SLC27A2 (107)
LIPC (108)
LIPC (109)
LACTB (110)
SCAPER (111)
ANPEP (112)
15
NUBP2 (113)
PDXDC1 (114)
ABCC1 (115)
ACSM5 (116)
ACSM2A (117)
ACSM2A (118)
SLC5A11 (119)
CETP (120)
GOT2 (121)
GOT2 (122)
SLC7A6 (123)
SLC7A5 (124)
16
RNF167 (125)
SLC13A5 (126)
SLC13A5 (127)
ALOX12 (128)
ACE (129)
CCDC57 (130)
17
LIPG (131)
18
GCDH (132)
ISYNA1 (133)
SLC7A9 (134)
APOE (135)
SULT2A1 (136)
FUT2 (137)
MBOAT7 (138)
19
SPTLC3 (139)
20
CBS (140)
21
PRODH (141)
COMT (142)
GGT1 (143)
TYMP (144)
ARSA (145)
22
Amino acid
Carbohydrate
Cofactor or vitamin
Energy
Lipid
Nucleotide
Peptide
Xenobiotic
Unknown


TwinsUK
N=6,056
503 Metabolites
KORA
N=1,768
517 Metabolites
GGM network
Primary discovery
486 overlapping metabolites in 7,824 participants
GWAS meta-analysis
(HapMap2 imputed, N=2.1M SNP)
137 lead SNPs
$P_{meta} \leq 1.03x10^{-10}$
$P_{het} \geq 0.001$
$P_{cohort} \leq 0.05/486$
Secondary discovery
98,346 metabolite ratios
log(m1 / m2) = log(m1) - log(m2)
GWAS meta-analysis
(HapMap2 imputed, N=2.1M SNPs)
8 lead SNPs
$P_{meta} \leq 5.08x10^{-13}$, P-gain > 250
$P_{het} \geq 0.001$
$P_{cohort} \leq 0.05/486$
145 lead SNPs
(299 SNPs-trait associations)
Locus annotation
• 1KGP association refinement
• Heritability estimates
• SNP*SNP interactions
• eQTLs causal analysis
Gene and disease annotation
• SNP-metabolite annotation
• Pathway gene enrichment
• Common disease
• Inborn error of metabolism
• Drug targets

Network of genetic loci and metabolic pathways

- Graph connects related pathways and loci associated with each
- Pairwise correlations of metabolites have been corrected from confounding effects (age, sex, etc)
- Stability of network measured by bootstrapping-based subsampling
  - Generated 1000 bootstrap datasets from original data and saw low variation

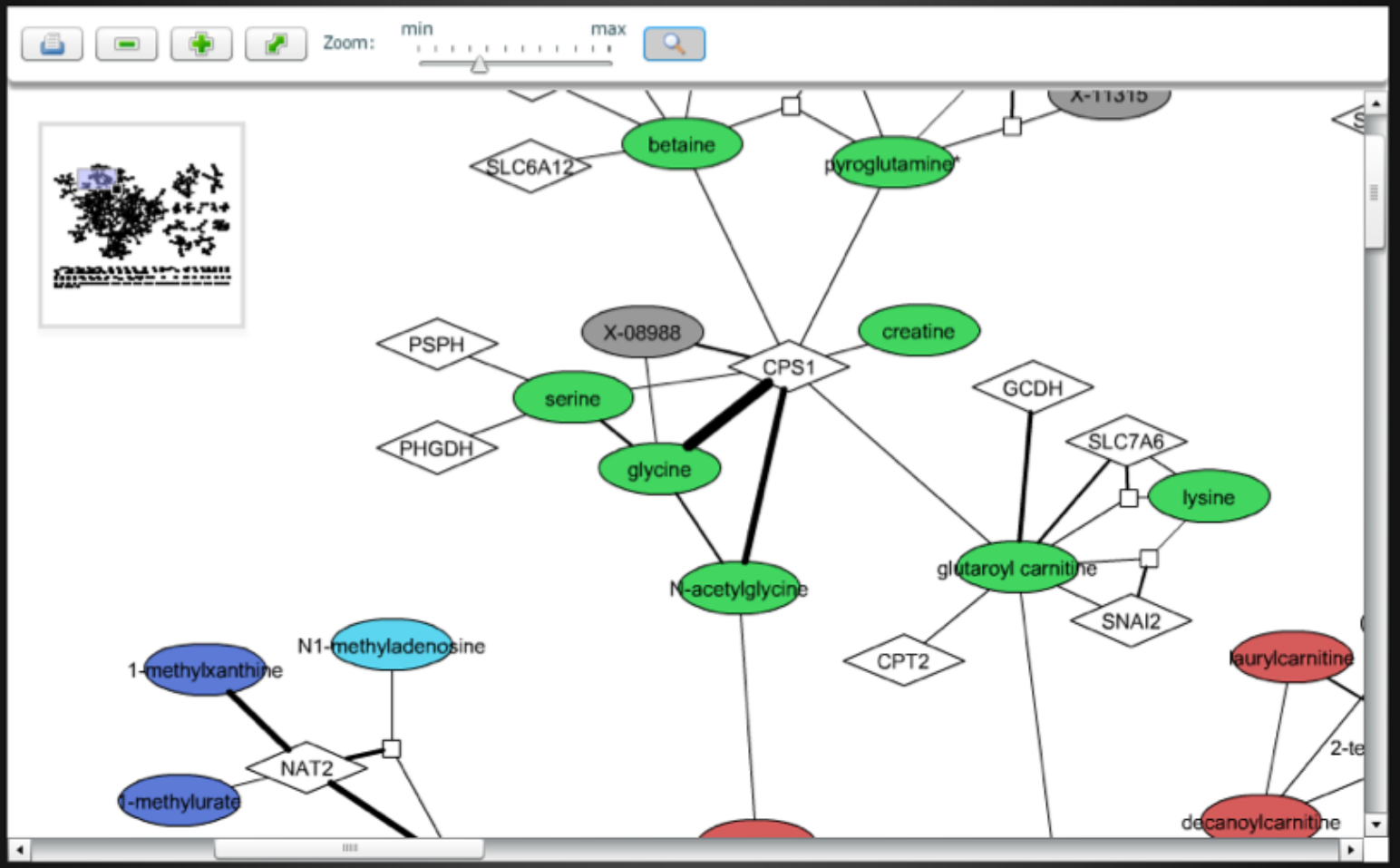

Metabolic pathways
1 Monoacylglycerol (4)
2 Lysolipid (23)
3 Glycerolipid (4)
4 Fatty acid–amide (1)
5 Fatty acid–branched chain (1)
6 Fatty acid–long chain (17)
7 Fatty acid–medium chain (9)
8 Fatty acid–short chain (1)
9 Fatty acid–dicarboxylate (5)
10 Fatty acid–essential (7)
11 Fatty acid–other (3)
12 Eicosanoid (1)
13 Carnitine (14)
14 Ketone body (1)
15 Bile acid (6)
16 Sterol or steroid (14)
17 Sphingolipid (1)
18 Inositol (2)
19 Benzoate (5)
20 Food component or plant (5)
21 Sugar or starch (1)
22 Xanthine (5)
23 Purine–adenine (1)
24 Purine–urate (2)
25 Purine–xanthine or inosine (4)
26 Purine–guanine (2)
27 Pyrimidine–uracil (2)
28 NAD metabolism (1)
29 Krebs cycle (4)
30 Oxidative phosphorylation (2)
31 Fructose, mannose and galactose (4)
32 Glycolysis and gluconeogenesis (5)
33 Aminosugar (1)
34 Creatine (2)
35 Glycine, serine and threonine (6)
36 Lysine (3)
37 Glutamate (3)
38 Alanine and aspartate (4)
39 Histidine (2)
40 Cysteine, methionine, SAM and taurine (4)
41 Butanoate (3)
42 Branched-chain amino acid (14)
43 Phenylalanine and tyrosine (10)
44 Glutathione (2)
45 Tryptophan (10)
46 Urea cycle, arginine and proline (8)
47 Guanidino and acetamido (1)
48 Amino fatty acid (1)
49 Polyamine metabolism (1)
50 Polypeptide (2)
51 γ-glutamyl (5)
52 Dipeptide (9)
53 Fibrinogen cleavage peptide (3)
54 Ascorbate and aldarate (3)
55 Tocopherol (2)
56 Hemoglobin and porphyrin (5)
57 Vitamin B6 (1)
58 Pantothenate and CoA (1)
Lipid
Carbohydrate
Xenobiotic
Amino acid
Nucleotide
Peptide
Energy
Cofactor or vitamin
XYZ New locus
XYZ Known locus
Genetic association
Metabolic association
Unknown
Association between metabolite ratio and locus
Phe-Leu
X-06307
Asp-Phe
Leu-Ala
X-11805
Phe-Ser
ACE
X-14086
Phe-Ala
α-Glu-Tyr
Leu-Ala
DSGEGDFXAEGGGVR
ABO
ADpSGEGDFXAEGGGVR
FUT2
ALPL
ADSGEGDFXAEGGGVR

Association map

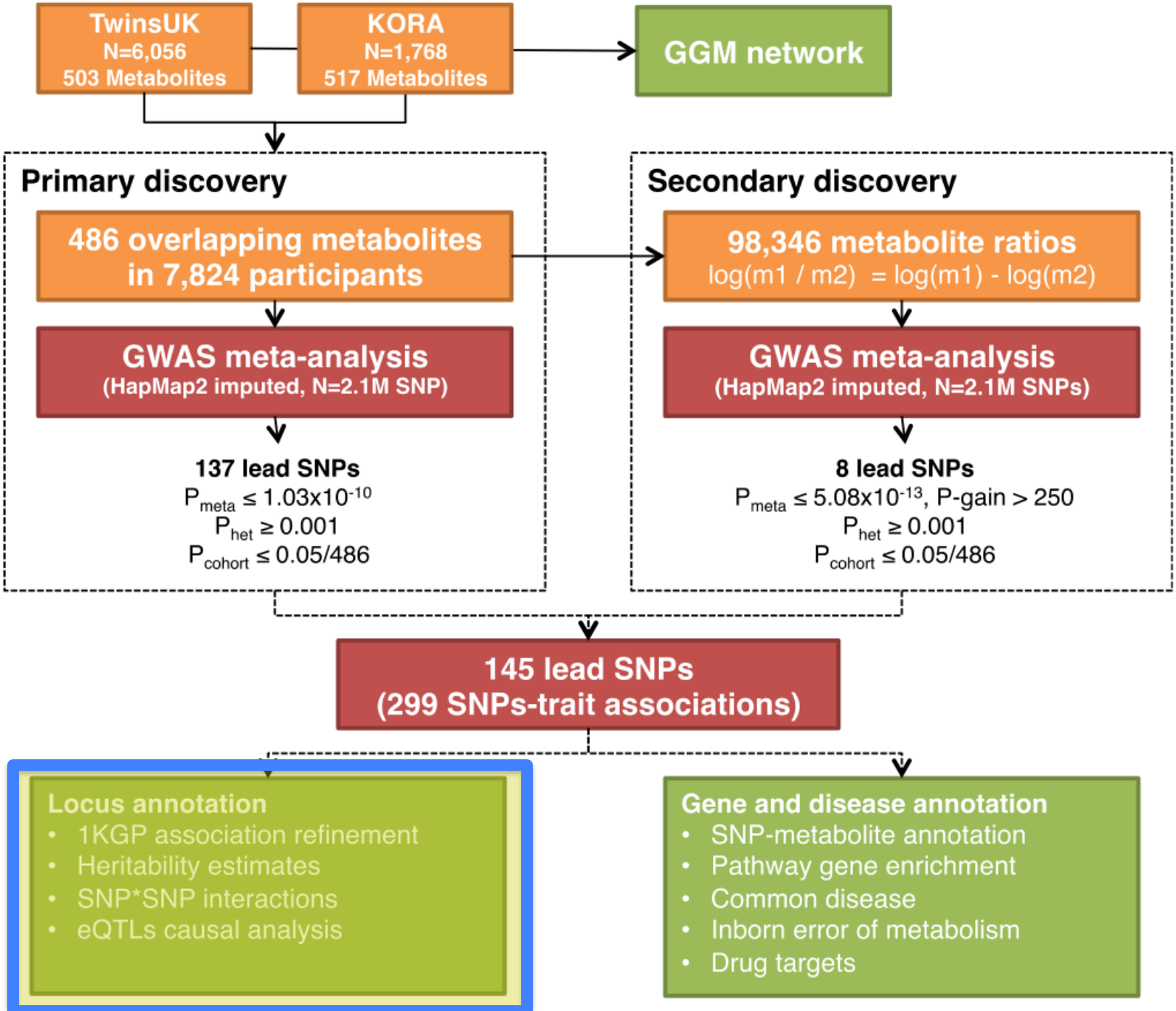
Network view: Locus 22

USAGE:

Zoom: Mouse wheel · Move graph: Left-click and hold while moving the mouse · Details: hover over nodes/edges · Navigate to locus: click on locus node



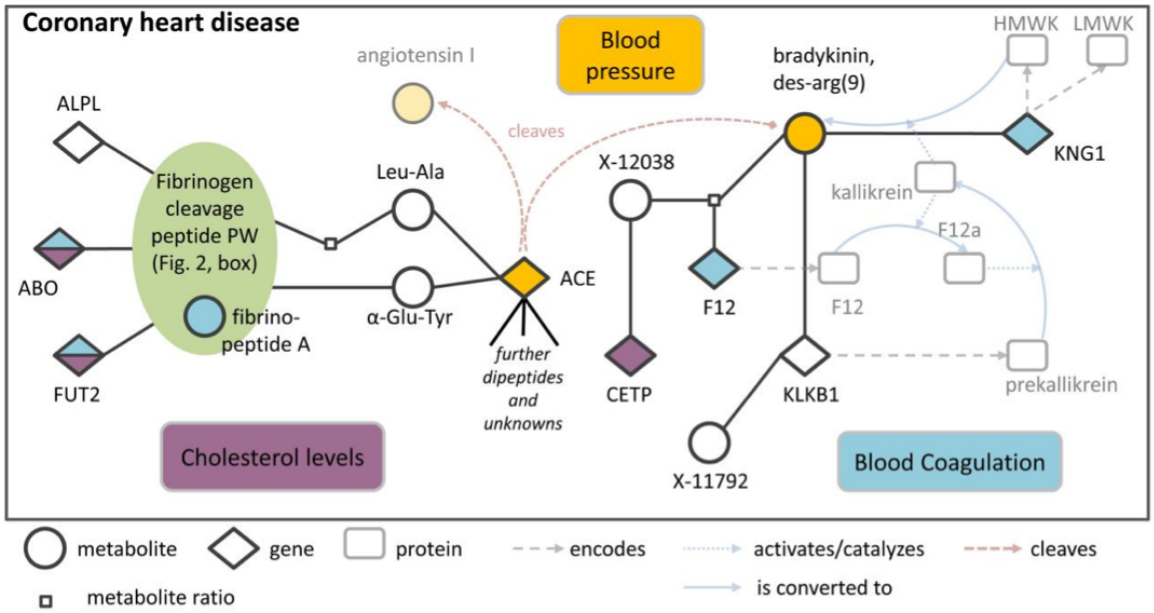
Download network in Cytoscape format


TwinsUK
N=6,056
503 Metabolites
KORA
N=1,768
517 Metabolites
GGM network
Primary discovery
486 overlapping metabolites in 7,824 participants
GWAS meta-analysis
(HapMap2 imputed, N=2.1M SNP)
137 lead SNPs
$P_{meta} \leq 1.03x10^{-10}$
$P_{het} \geq 0.001$
$P_{cohort} \leq 0.05/486$
Secondary discovery
98,346 metabolite ratios
log(m1 / m2) = log(m1) - log(m2)
GWAS meta-analysis
(HapMap2 imputed, N=2.1M SNPs)
8 lead SNPs
$P_{meta} \leq 5.08x10^{-13}$, P-gain > 250
$P_{het} \geq 0.001$
$P_{cohort} \leq 0.05/486$
145 lead SNPs
(299 SNPs-trait associations)
Locus annotation
• 1KGP association refinement
• Heritability estimates
• SNP*SNP interactions
• eQTLs causal analysis
Gene and disease annotation
• SNP-metabolite annotation
• Pathway gene enrichment
• Common disease
• Inborn error of metabolism
• Drug targets

Heritability and epistasis

- Estimates contribution of genetic vs. environmental factors to metabolite level variance
- For all metabolic loci:
  - Median 6.9% heritability (high)
- Low epistasis found (i.e. the loci act in an additive manner)

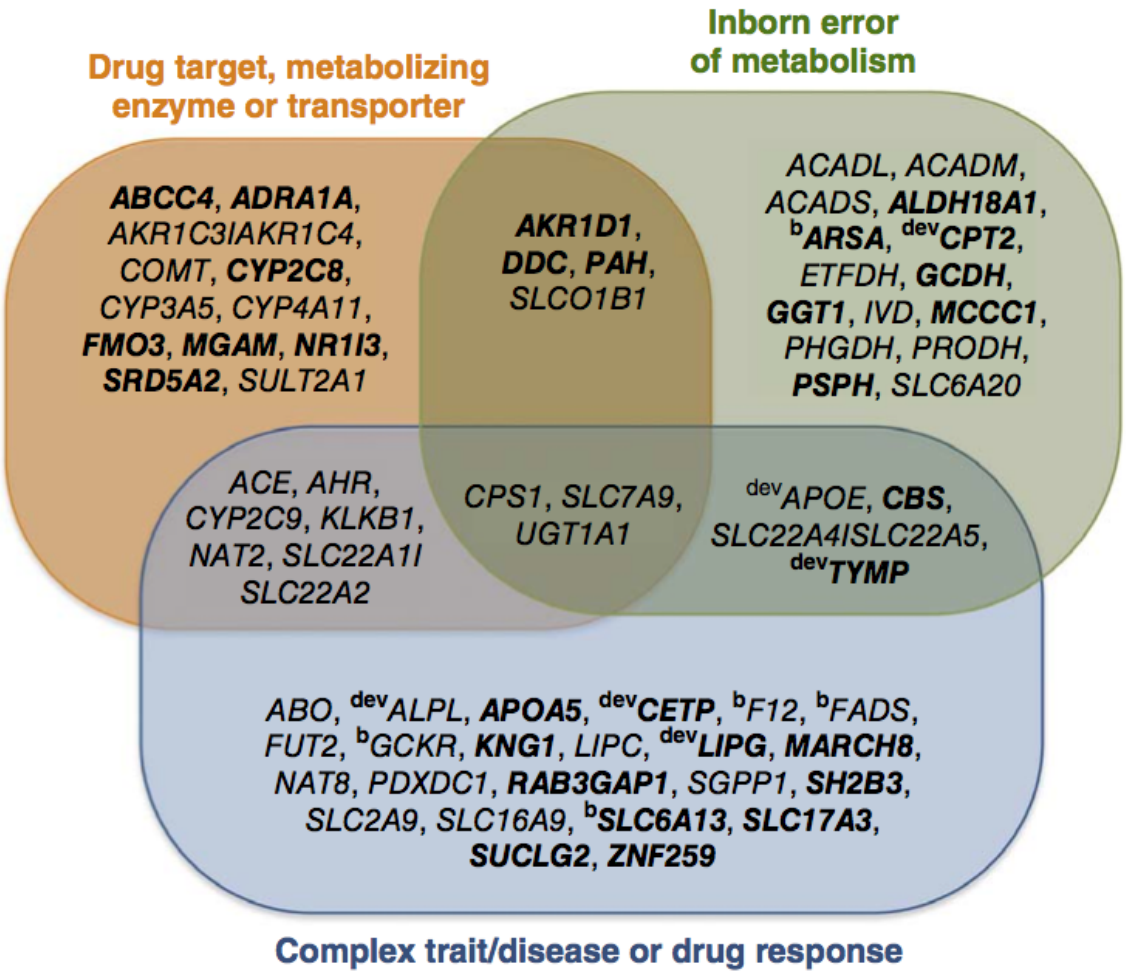
Using the network to infer biology



- Bradykinin-kininogen-kinin system regulates blood pressure and coagulation
- This subnetwork is from the atlas in created in this study
- Novel associations found between KNG1, F12 variants and bradykinin

Integrating associations with complex traits/disease

- Searched National Human Genome Rsearch Institute (NHGRI) GWAS catalog
  - Database contains information on SNP-disease associations
- Plotted the loci based on medical/ pharmacological significance



Enzyme

bABP1, ACMSD, ACSM2A, ACSM5, bADHFE1, AGXT, bALOX12, ANPEP, ASPG, bBHMT, bCCBL1, CRAT, ECI2, ELOVL2, GADL1, GBA3, GLS2, GMPR, GOT2, bHAO2, bIDO1, ISYNA1, bNT5E, OPLAH, PAOX, PPM1K, RNF167, devSCD, SKIV2L, SPTLC3, bTDO2, THEM4, devTKT, UGT3A1

Transporter

devABCC1, ANKH, bOSTalpha, SLC5A6, devSLC7A5, SLC13A5, SLC16A10, bSLC1A4, bSLC27A2, bSLC5A11, bSLC6A12, SLC7A6

Growth factor

ANGPTL3

Ion channel

PEX5L

Transcriptional regulator

ALX3, DMBX1, SNAI2

Other

CCDC57, DIP2C, HEATR4, HPS5, bLACTB, MBOAT7, NASP, NUBP2, PM20D2, PRRC2A, PYROXD2, SCAPER, WDR66, ZCWPW1

Bold = New associations found in this study
"dev" = Genes associated with compounds in drug development
"b" = Bioactivity

Summary

- Found genome-wide associations at 145 metabolic loci connected with 400 metabolites in human blood
- Characterized heritability, epistatic effects
- Characterized overlaps with known loci of complex disorders, inborn errors of metabolism, and drug targets
- Developed database for data mining and visualization

Medical importance

- Newly identified loci can be related with genes involved in disorders, complex traits, and drug response
- Discovery of new drug targets
- Serum and urine metabolites as biomarkers
  - Assess patient response to a drug
  - Taylor new treatment/personal medicine

Questions?