

Metabolomics in Sickle Cell Disease: Searching for severity biomarkers

Yann Ilboudo

MHIGFG meet January 17th 2018

Outline

- Introduction to Sickle Cell Disease (SCD)
- Overview of Metabolomics Experiments
- Data Analysis: Targeted vs. Untargeted
- Conclusions & Perspectives

Sickle Cell Disease – Genetic Mutation & Consequences

Beta-globin gene

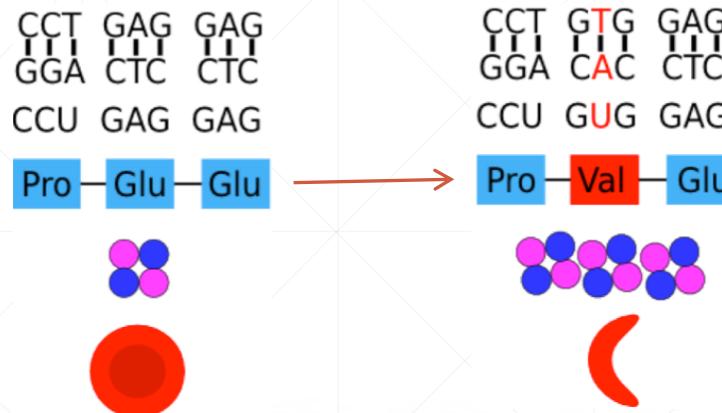
CCT GAG GAG
GGA CTC CTC
CCU GAG GAG

Pro — Glu — Glu



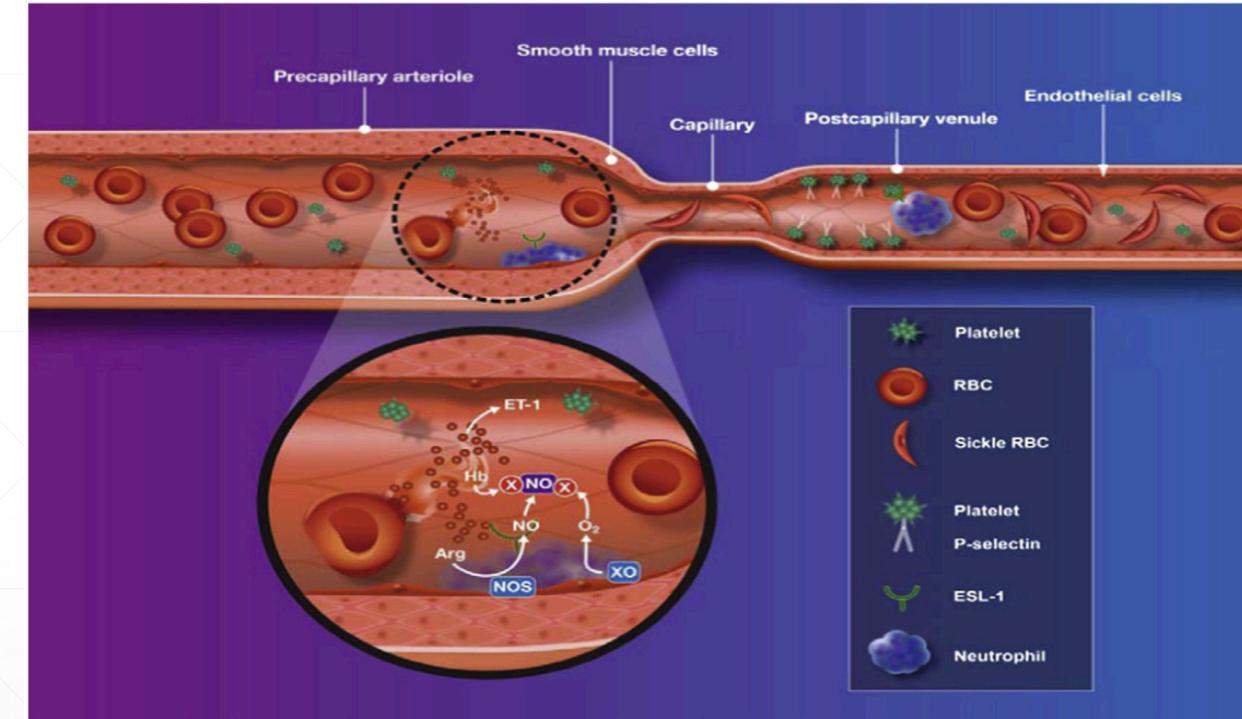
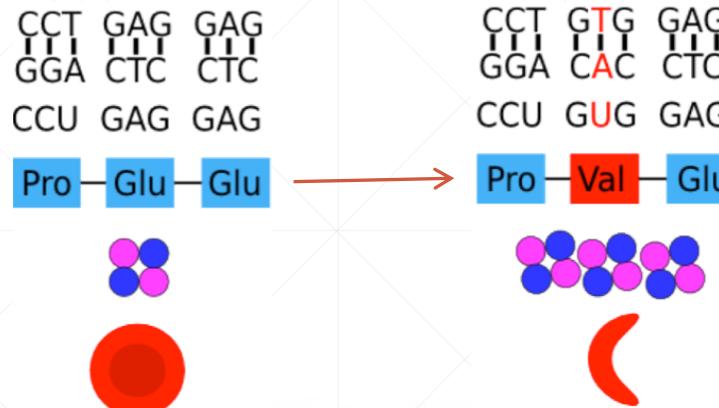
Sickle Cell Disease – Genetic Mutation & Consequences

Beta-globin gene



Sickle Cell Disease – Genetic Mutation & Consequences

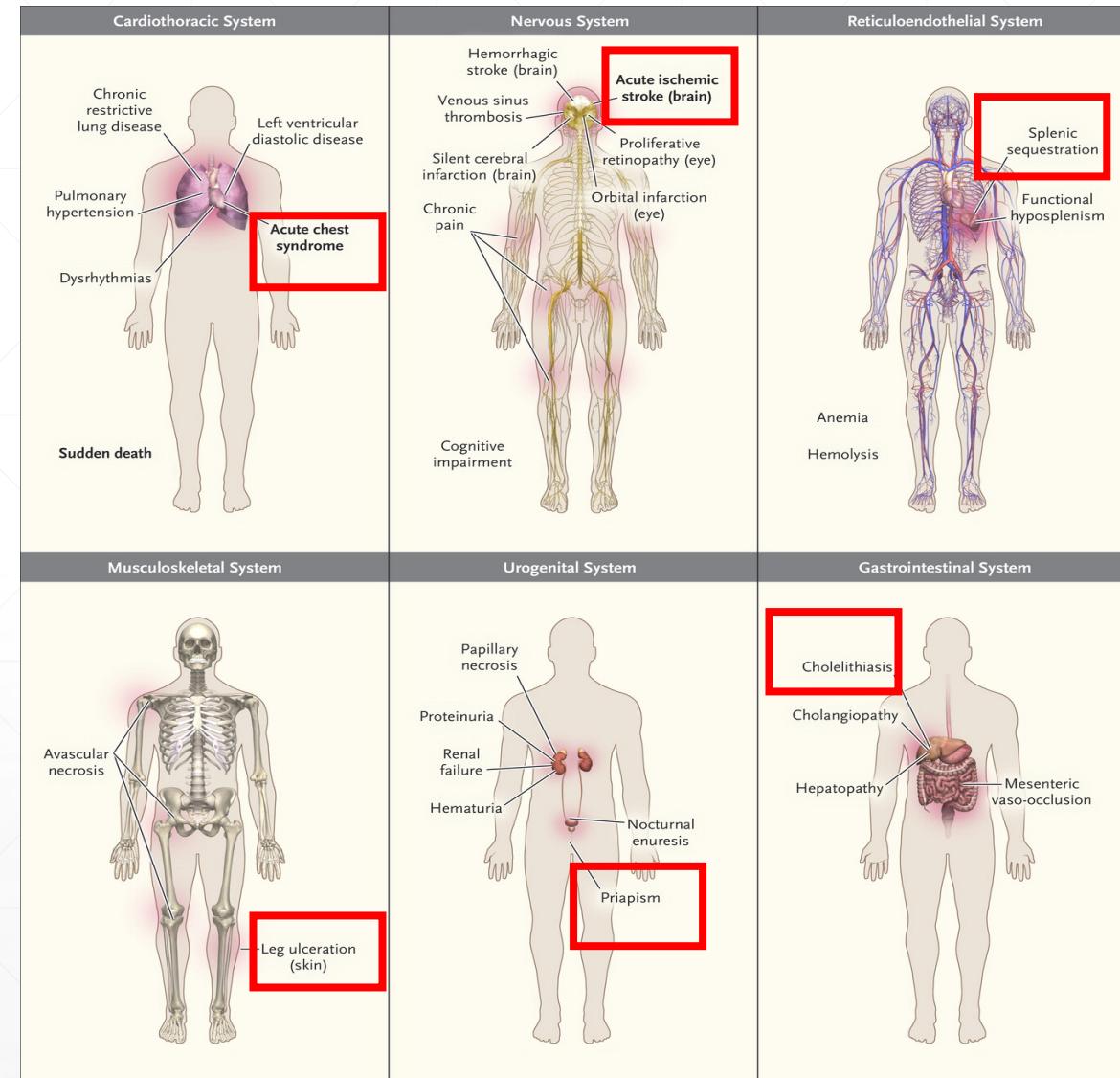
Beta-globin gene



- Dehydrated red blood cell
- Sickle red cell stick to the endothelial walls
- Vaso-constriction of capillaries

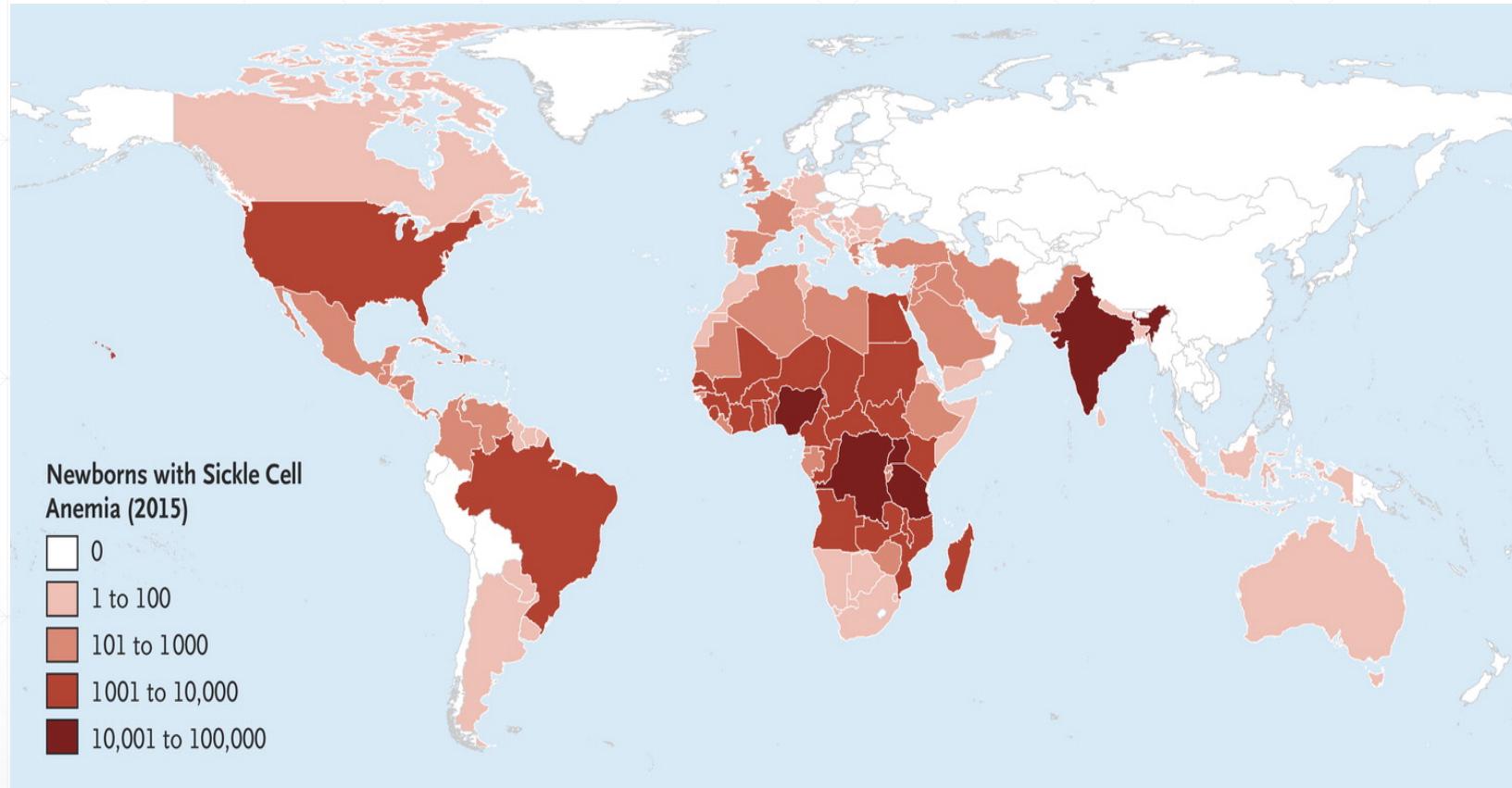
Sickle Cell Disease – Complications

- Systemic disease, affecting multiple organ systems.
 - Respiratory system
 - Acute chest syndrome
 - Digestive system
 - Liver disease
 - Gallbladder dysfunction
 - Urogenital system
 - Nephropathy
 - Priapism
 - Nervous system
 - Stroke
 - Musculoskeletal system
 - Leg ulcer



SCD Burden

- Estimate suggest that there are ~300,000 newborn every year with SCD.
- Prevalence is highest in sub-Saharan, and central Africa, as well as in India. Incidence is also important in the Mediterranean basin, the Middle East and the Americas.



Sickle Cell Disease Treatment

- Hydroxyurea treatment – FDA approved
 - Its goal is to prevent the polymerization of abnormal hemoglobin
 - The response amongst patients is wide, and taking has many negative side effects
- Endari (L-glutamate) treatment – FDA approved (2016)
 - First treatment approved in the last 20 years
 - Reduces oxidative stress and acute complications



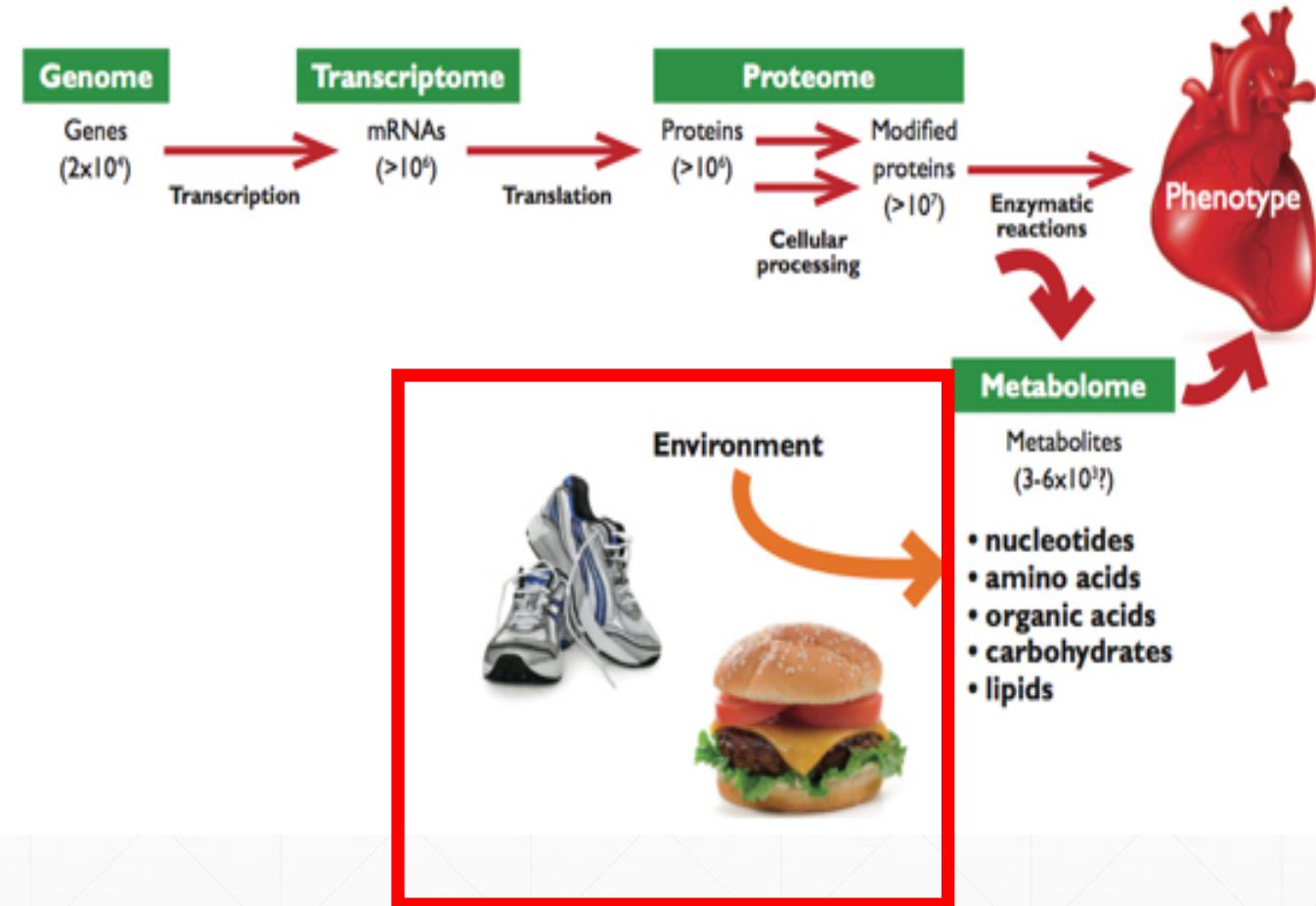
Goals and hypothesis of this project

- Hypothesis 1: A combination of both genetic variations and metabolic signatures are more accurate descriptors of SCD severity.
 - Goal 1: Identify metabolites that correlate with blood traits (e.i. MCV, MCH) and SCD related complications.
 - Goal 2: Identify metabolites associated with genetic variants.
 - Goal 3: Establish causal link between metabolites and SCD-related traits (blood/complications) through genetics.
- Hypothesis 2: Untargeted metabolomics approach can enable discovery of SCD biomarkers.
 - Goal 1: Identify changes in metabolome in SCD patients for complications.
 - Goal 2: Relate these changes to pathways.

What is Metabolomics?

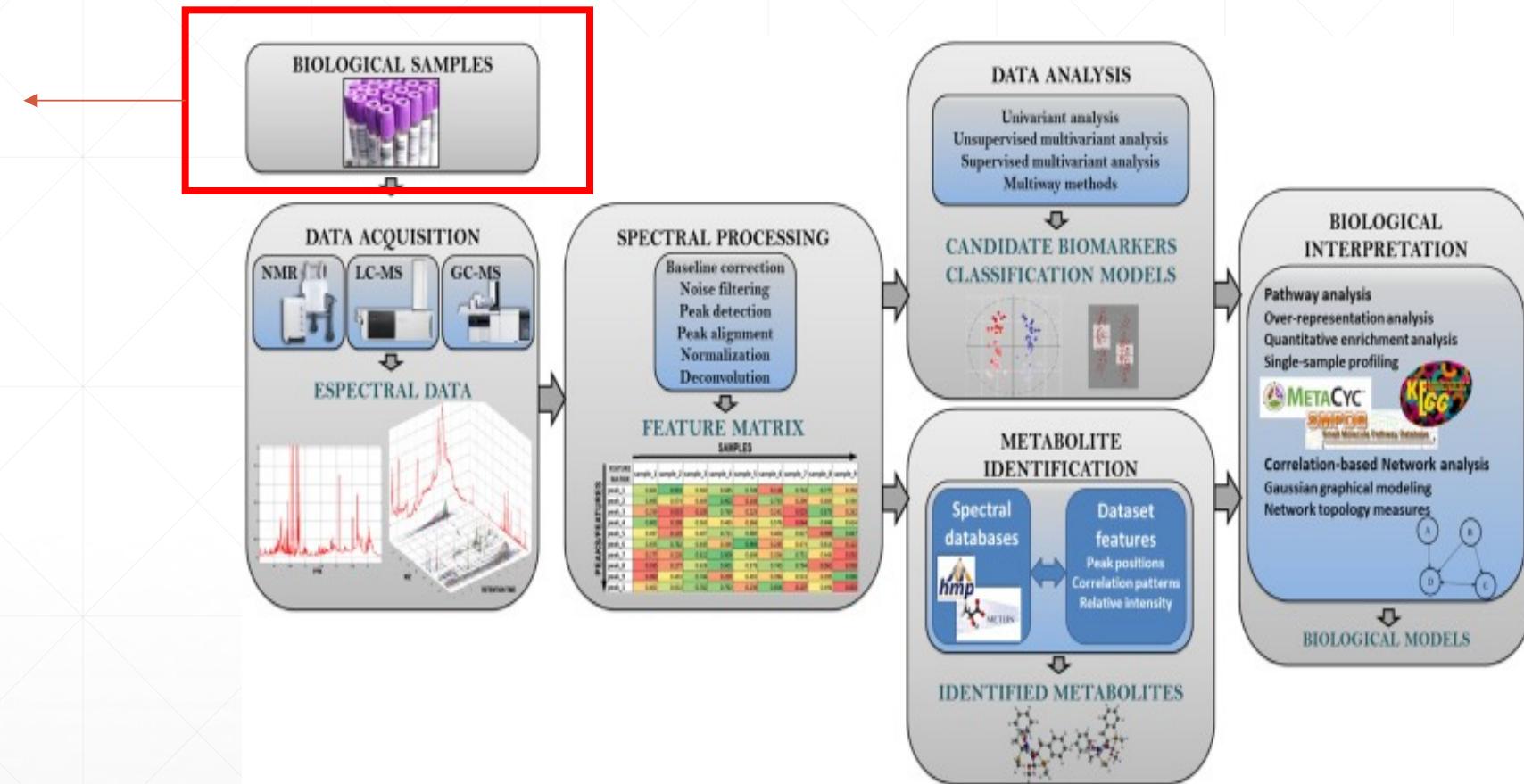
Introduction

- Metabolomics is the method for profiling and analysis metabolites.
- Metabolites represent the integration of gene expression, protein interaction and other different regulatory processes and the environment.
- Ideal: For understanding and tracking by-products of physiological processes relevant to diseases, lifestyle choices, medications.



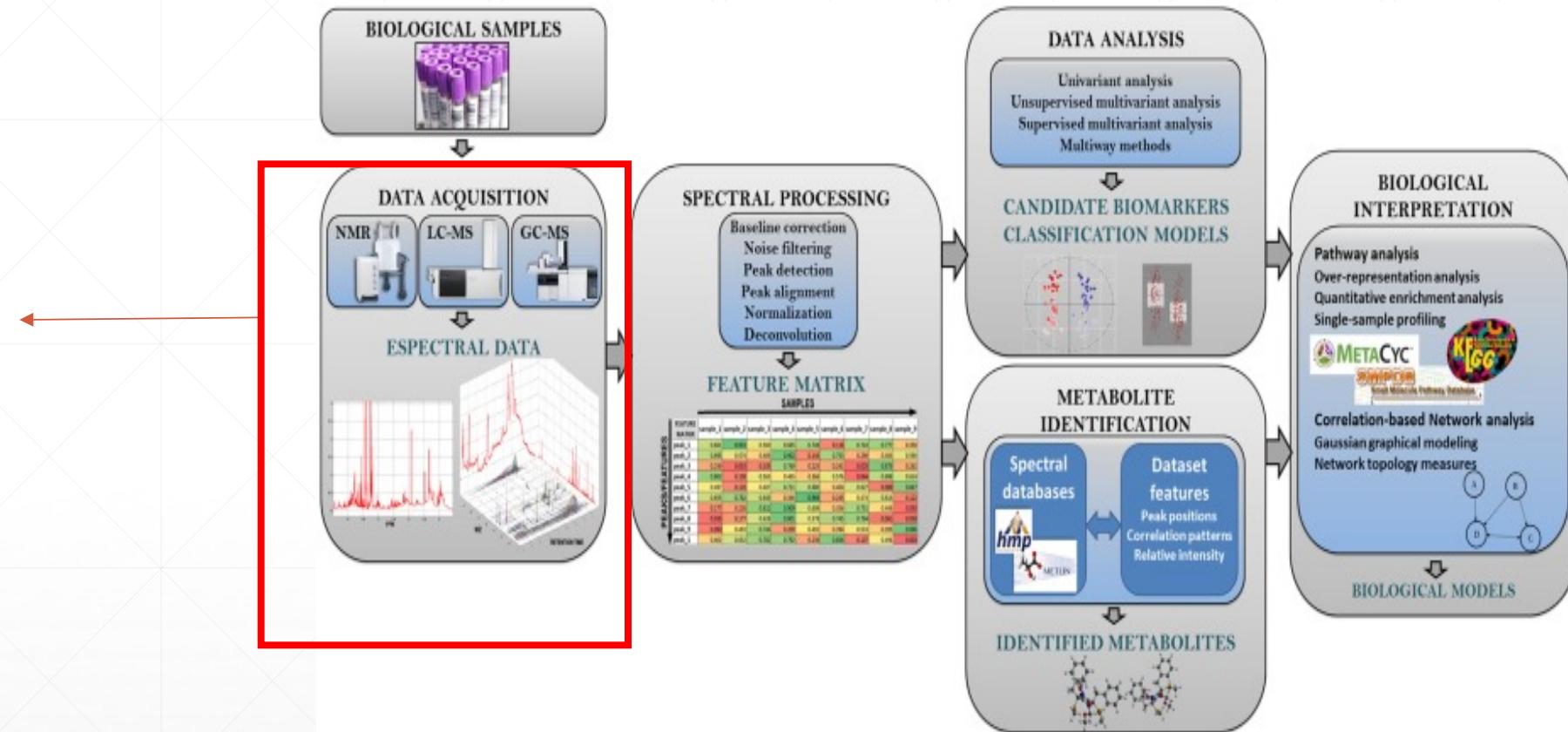
Overview of Metabolomics – Typical Pipeline

Sample from
blood, urine, saliva
or other tissue



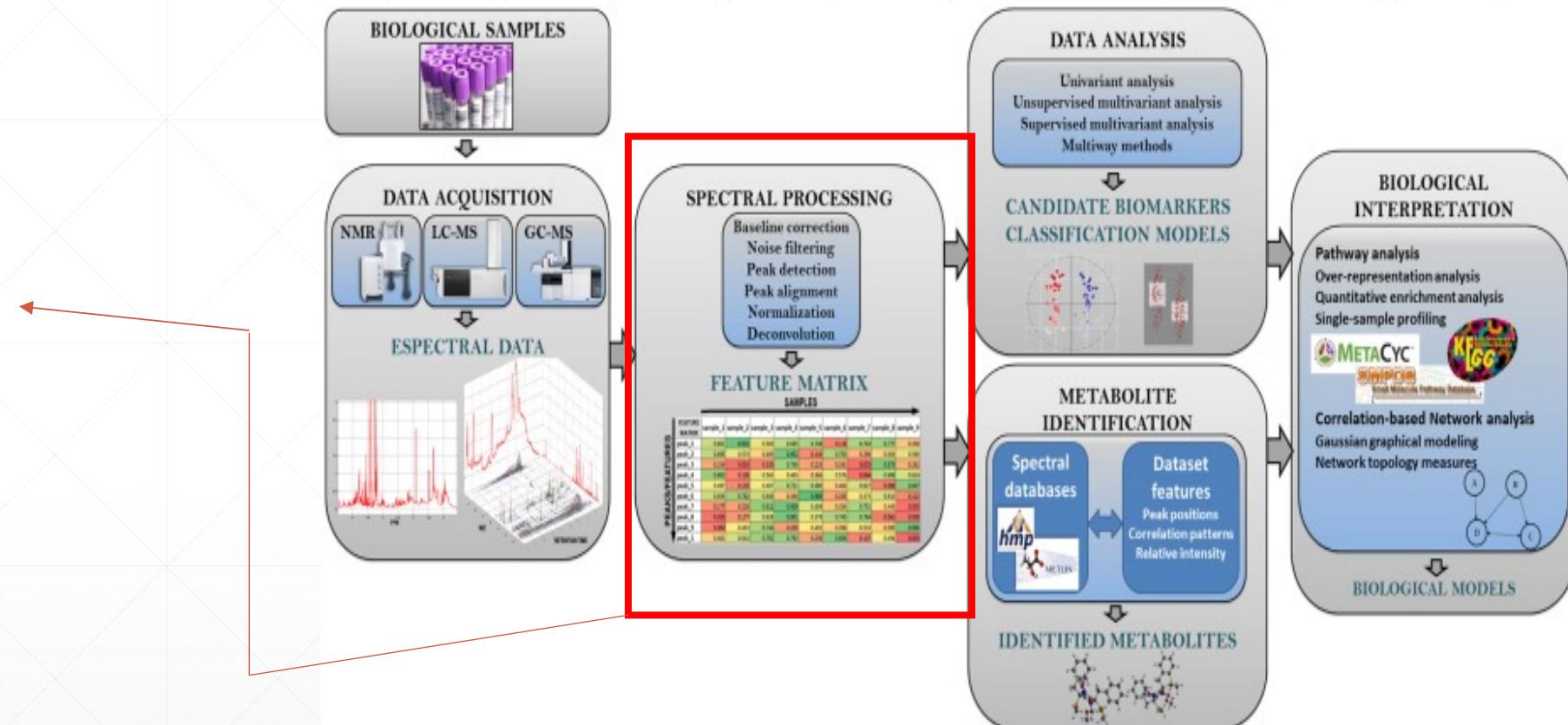
Overview of Metabolomics – Typical Pipeline

- Two main methods:
 - Nuclear Magnetic Resonance (NMR) spectrometry.
 - MS based



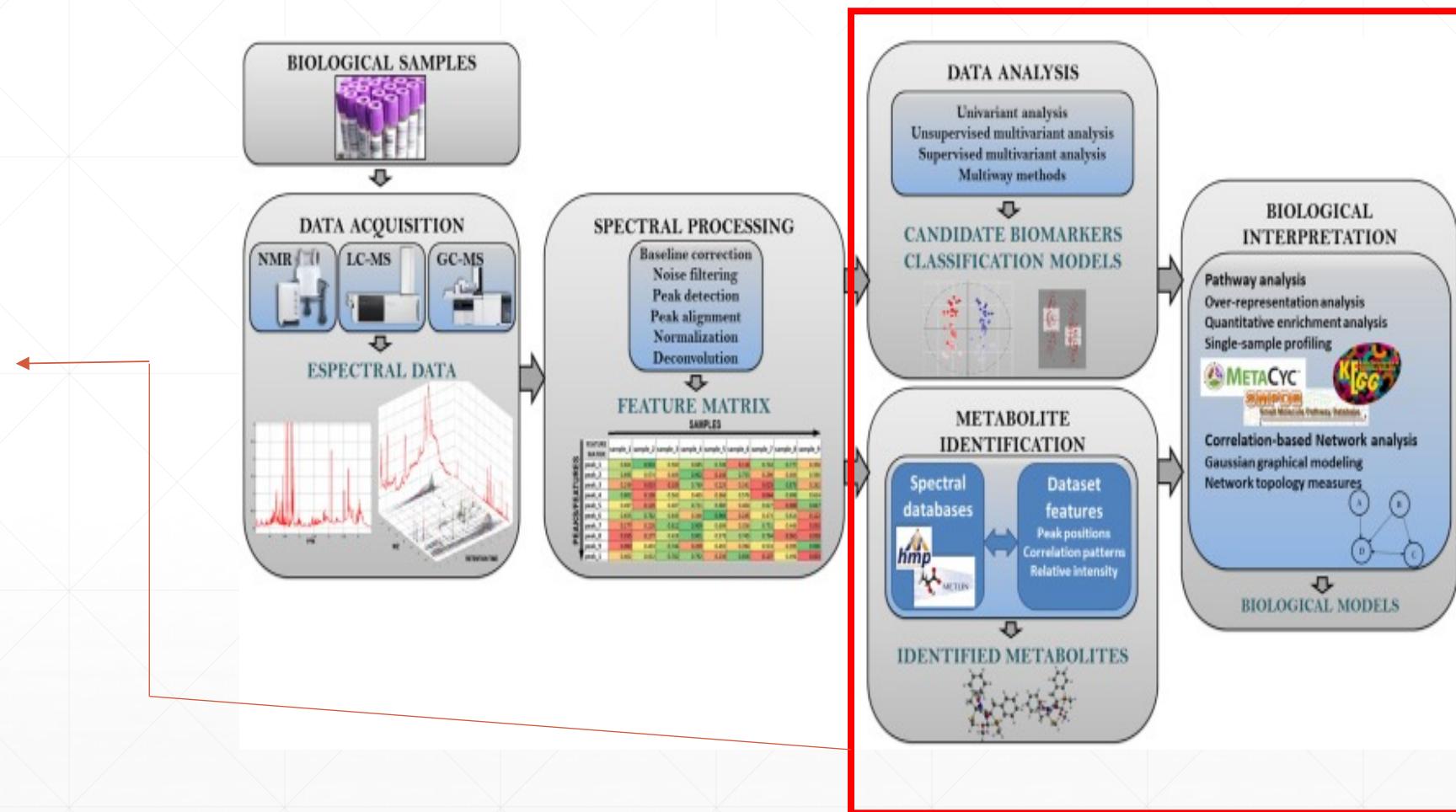
Overview of Metabolomics – Typical Pipeline

- Metabolite Identification
- QC (noise filtering, peak detection, peak deconvolution, retention time alignment)



Overview of Metabolomics – Typical Pipeline

- Data Analysis
 - For example: Regressions, principal component analysis (PCA), partial least-square discriminant (PLS-DA)



Overview of Metabolomics – Targeted vs. Untargeted

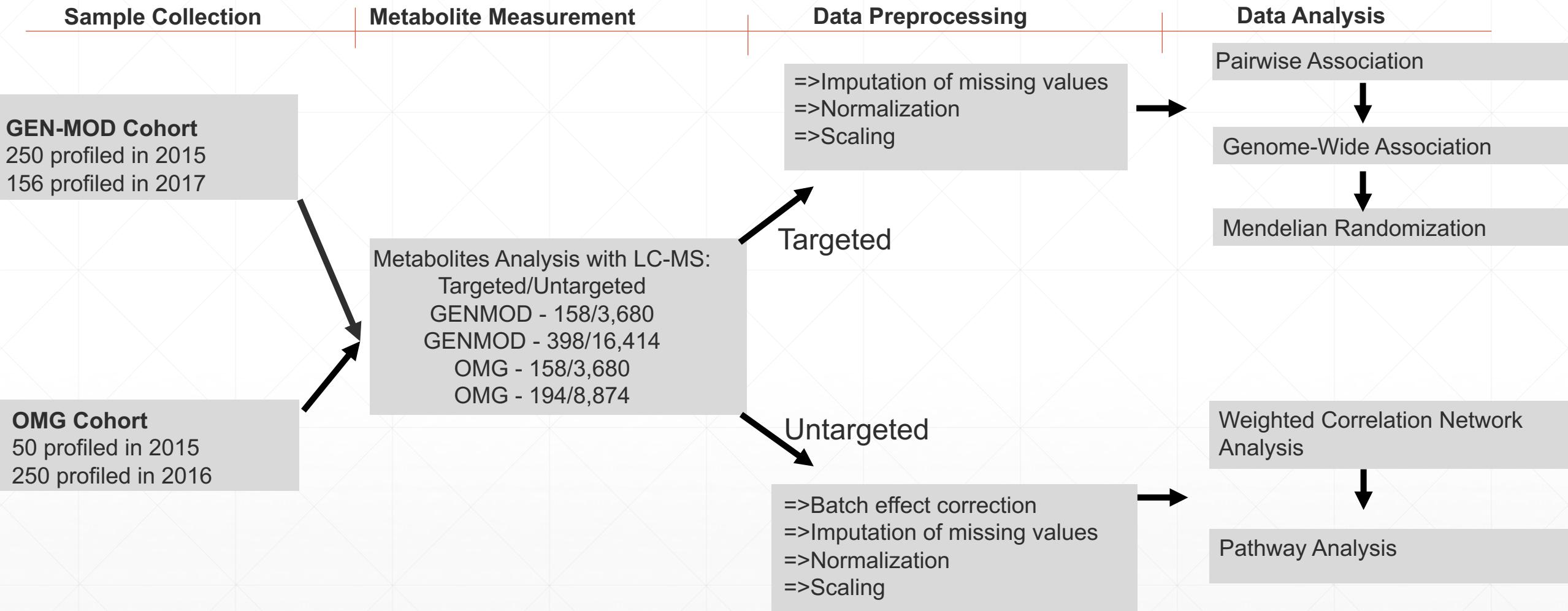


Yields: ~100s metabolites
Hypothesis driven
Comparative studies

Yields: ~10,000s metabolites
Hypothesis generating
Comparative studies
Allow discoveries of new pathways

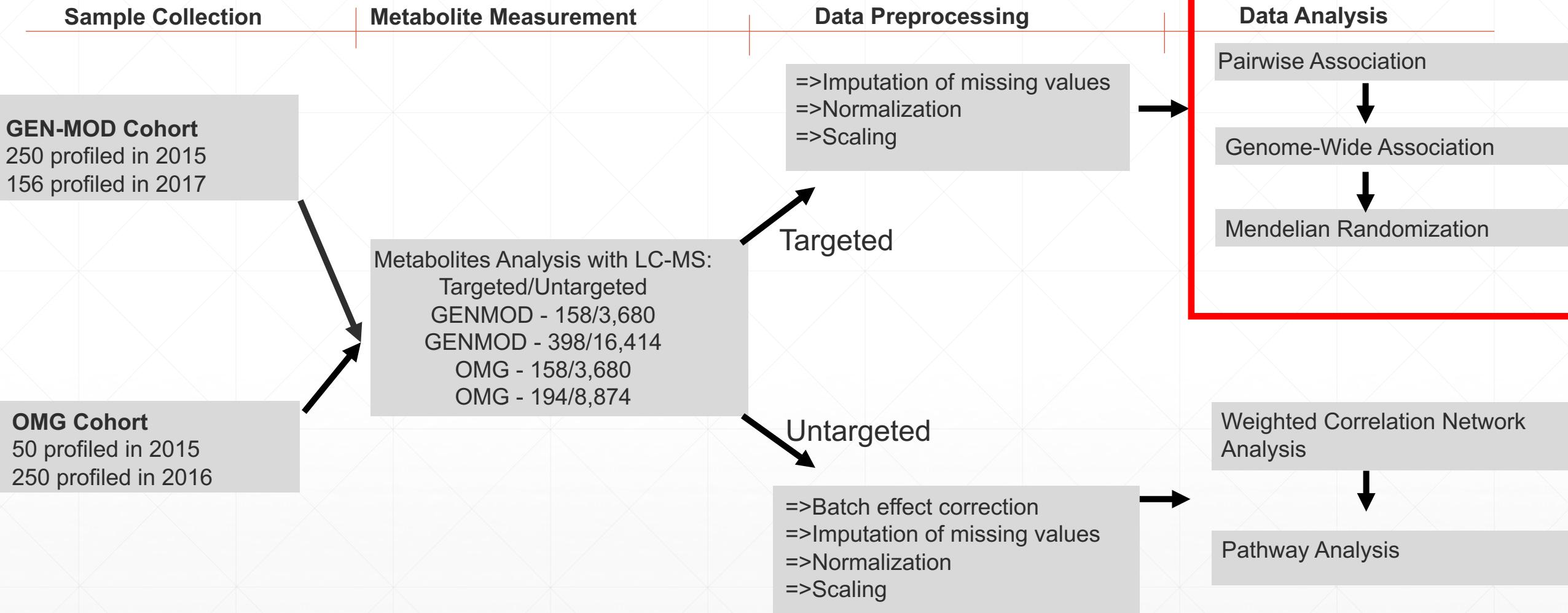
SCD Study Design Metabolomics

Methods



SCD Study Design Metabolomics

Methods



Data Analysis – Targeted Approach

Pairwise Association

1



Genome-Wide Association



Mendelian Randomization

Data Analysis – Pairwise Analysis - Targeted

- Reminder Goal 1: Identify metabolites that correlate with blood traits (e.i. MCV, MCH) and SCD related complications (e.i, leg ulcers, retinopathy).
 - Test 129 known metabolites against 11 blood traits and 8 SCD complications in N=706 patients.
 - Out of these, 81 metabolites showed strong associations (Bonferroni correction: $P\text{-value} < 2.4 \times 10^{-5}$).
 - 1 metabolites with complications.
 - 80 metabolites with blood traits.

Data Analysis – Pairwise Analysis – Targeted – Bile Acids

Results

- Reminder: Bile acids are formed from cholesterol
- SCD patients go through gallbladder removal because they end with gallstone.
- This results shows that glycodeoxycholate is strong indicator of liver dysfunction in SCD, or of medication to reduce gallstones.

Families	Metabolites	Cholecystectomy
Bile acids and derivatives	Glycocholate	1.2*
Bile acids and derivatives	Glycodeoxycholate/Glycochenodeoxycholate	1.53****
Bile acids and derivatives	Lithocholate	1.29*

Data Analysis – Pairwise Analysis – Targeted – Bile Acids

Results

- Reminder: Bile acids are formed from cholesterol
- Lithocholate association with reduction of RBC count, and hematocrit.
- This result is in agreement with that there is a dysregulation of bile acids in SCD probably mediated through hemolysis.

Families	Metabolites	RBC	Hematocrit
Bile acids and derivatives	Glycocholate	-0.0814*	-0.1047*
Bile acids and derivatives	Glycodeoxycholate/Glycochenodeoxycholate	-0.1226***	-0.1307***
Bile acids and derivatives	Lithocholate	-0.3127*****	-0.2835*****

Data Analysis – Targeted Approach

Pairwise Association



Genome-Wide Association



Mendelian Randomization

Data Analysis – GWAS - Targeted

- Reminder Goal 2: Identify metabolites associated with genetic variants.
 - Test 81 metabolites against 30M markers in 706 SCD patients.
 - Results for sample size of 401, as we are waiting for associations results from OMG group to meta-analyze individuals with GENMOD.
 - Selected variants with MAF > 1%, and *P-value* < 1×10^{-8} .
 - Results in 9 metabolites remained.
 - ADMA/Arginine_derivatives
 - Lithocholate/Bile Acid
 - AMP/Purines_and_derivatives
 - N.carbamoyl.beta.alanine/Urea_and_derivatives
 - Dimethylglycine/ Glycine_derivatives
 - Carnitine/ Acyl_carnitines
 - C3.carnitine/ Acyl_carnitines
 - C26 carnitine/Acyl_carnitines
 - Aminoisobutyric.acid/ Gamma-amino_acids

GWAS – Results – Lithocholate/Bile Acids

- UGT1A4 is a well known polymorphism in SCD and is associated with bilirubin levels
- Large GWAS, ~ 8,000 healthy Europeans found those SNPs to be associated with bilirubin or biliverdin.

MarkerName	Freq1	Pvalue	Direction	Consequence	Existing_variation	SYMBOL
2_234668570_C_T	0.4825	2.36E-17	++	intron_variant	rs887829	UGT1A4
2_234672639_G_T	0.4578	1.97E-16	++	intron_variant	rs6742078	UGT1A4
2_234601669_T_G	0.6542	3.43E-08	--	missense_variant	rs6759892	UGT1A6

An atlas of genetic influences on human blood metabolites. So-Youn Shin, et al. Nature Genetics. 46, 543–550 (2014).

GWAS – Results – Carnitine/C3.carnitines

- SLC16A9 encodes a carnitine efflux transporter.
- Large GWAS, ~ 8,000 healthy Europeans found those SNPs to be associated with bilirubin or biliverdin.

Carnitines

MarkerName	Freq1	Pvalue	Direction	Consequence	Existing_variation	SYMBOL
10_61469090_C_T	0.7594	1.40E-12	++	intron_variant	rs1171615	SLC16A9

C3.carnitines

MarkerName	Freq1	Pvalue	Direction	Consequence	Existing_variation	SYMBOL
10_61469090_C_T	0.7595	4.33E-10	++	intron_variant	rs1171615	SLC16A9

An atlas of genetic influences on human blood metabolites. So-Youn Shin, et al. Nature Genetics. 46, 543–550 (2014).

GWAS – Results - Arginine

- Asymmetric dimethylarginine (ADMA) methylated arginine, has been linked to kidney disease.
- Association in the hemoglobin sub-unit alpha 2.

MarkerName	Freq1	Pvalue	Direction	Consequence	Existing_variation	SYMBOL
16_223678_C_<CN0>	0.2308	2.01E-09	--	3_prime_UTR_variant	-	HBA2

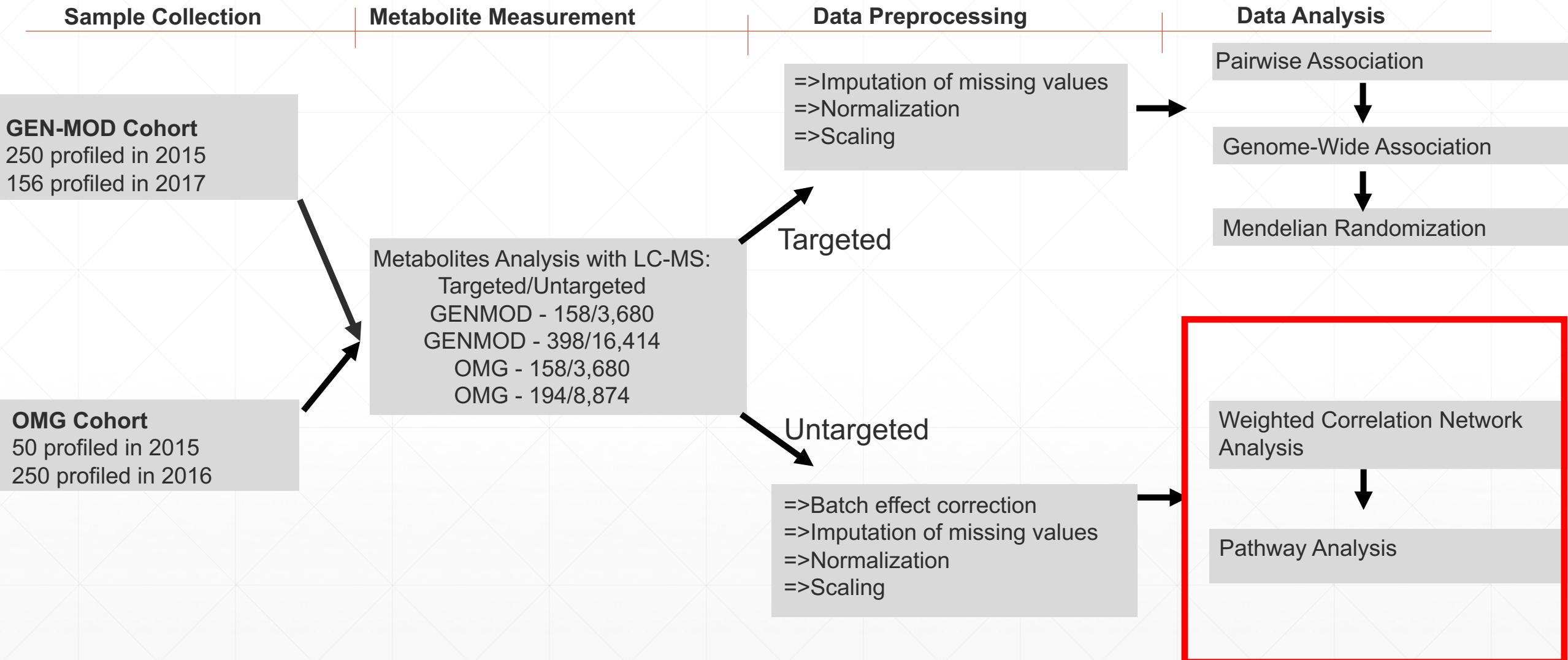
Data Analysis – Pairwise Association & GWAS - Conclusion

Conclusion

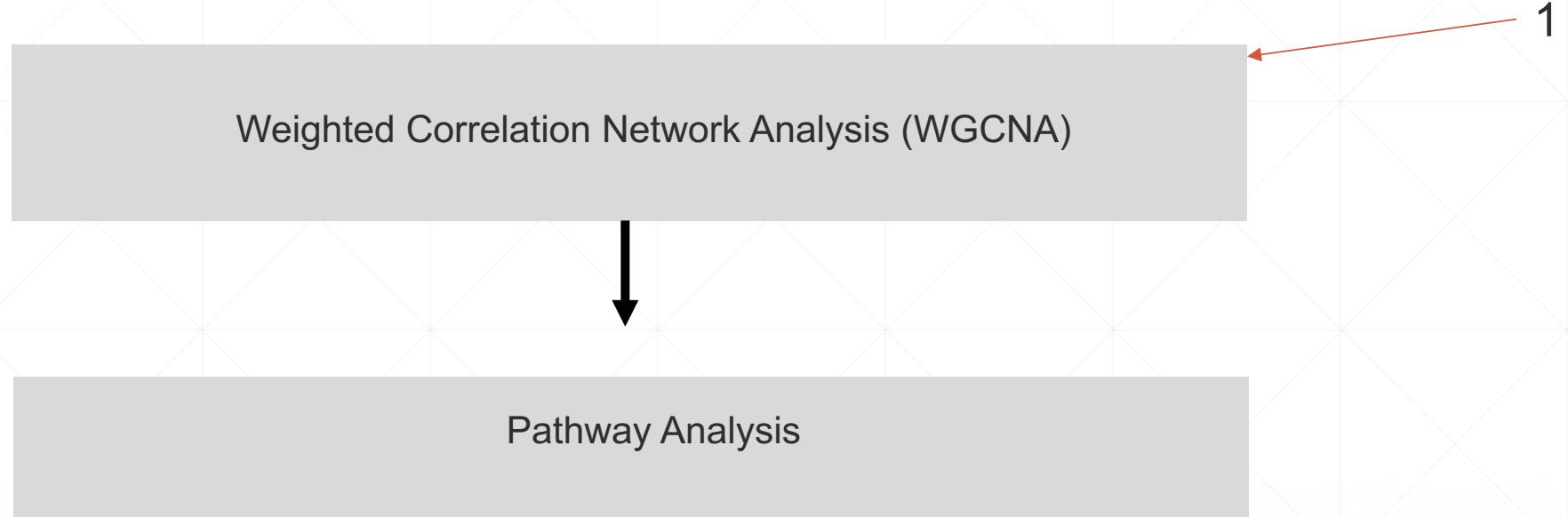
- We started our analysis with 129 metabolites
- Pairwise association highlighted 81 metabolites with strong associations
 - Showed relationship between bile acids metabolites and SCD diseases.
- Preliminary GWAS highlighted 9 metabolites
 - Found some new associations with phenotypes, and some already described associations.
- Next step Mendelian randomization, to build a case for causality between SNPs, disease outcome, and metabolites.
- Next step compare association found in SCD patients, with association of the same metabolites in healthy African-Americans.

Data Analysis

Methods

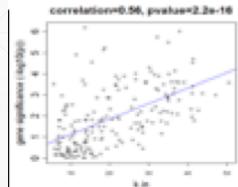
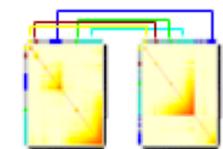
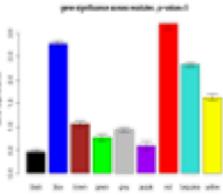
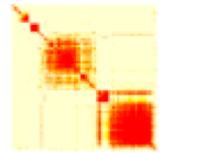


Data Analysis – Untargeted Approach



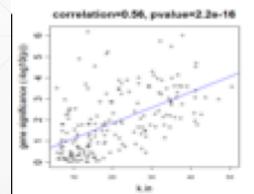
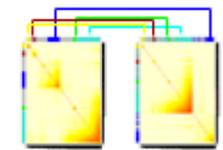
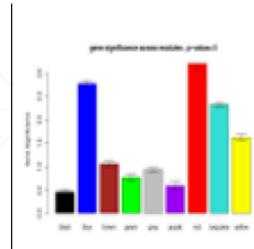
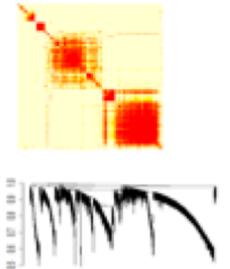
Data Analysis – Untargeted Approach - WGCNA

- WGCNA is a network based method that was originally developed for transcriptomics. Yet, it is applicable to other settings, such as metabolomics, cancer, and analysis of brain imaging data.
- We applied three steps to our analysis:
 - Construction of metabolite networks
 - Computing the correlation of all the metabolites.
 - Calculates the correlation of two metabolite taking into consideration the correlation of all other metabolite, as known as the Topological Overlap Map (TOM).
 - Then groups together metabolite that are similar to one another. (Hierarchical clustering)
 - Relating this networks to disease complications
 - Run associations between modules and trait of interest



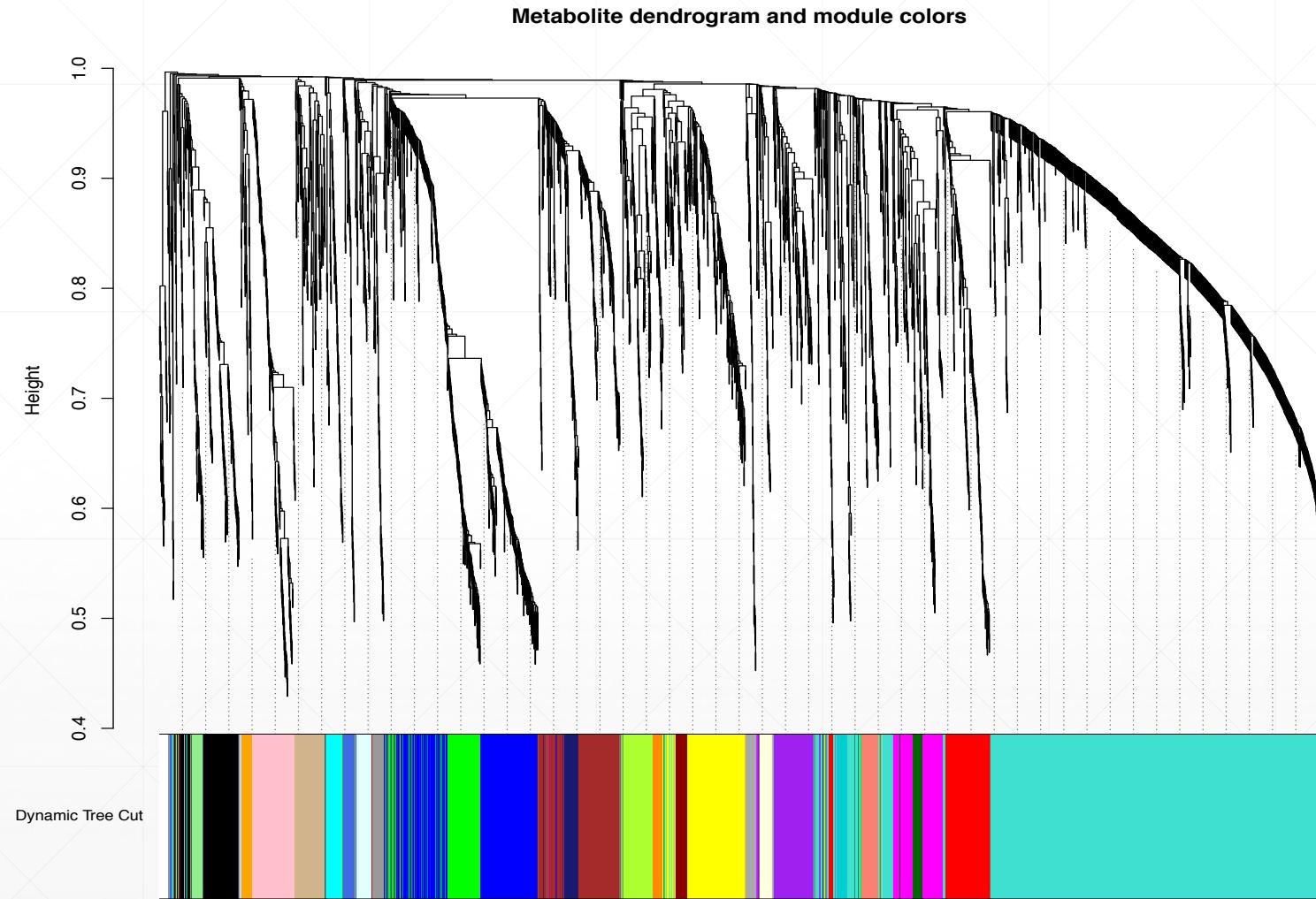
Data Analysis – Untargeted Approach - WGCNA

- Advantages to this approach:
 - No worries about multiple testing.
 - Study metabolites together as opposed to one at a time.
 - Understand how metabolites relate to each other inside give module.



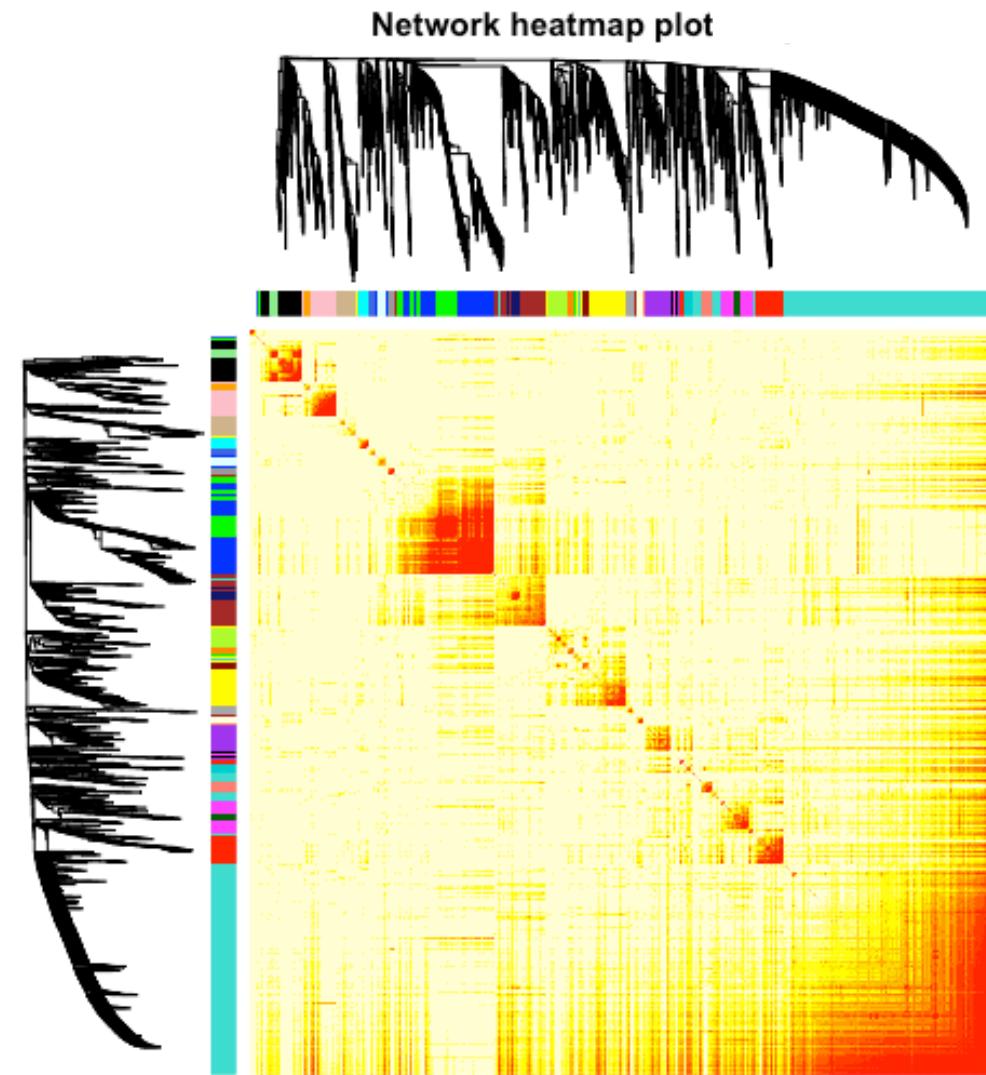
Data Analysis – Untargeted Approach – WGCNA - Results

- We applied WGCNA to the batch containing of 300 samples profiled in 2015 and about 4,000 metabolites.
- We got 28 networks with relatively large minimum size of 30 metabolites per networks.



Data Analysis – Untargeted Approach – WGCNA - Results

Legulcers		
Mebblue.p	Medarkgreen.p	Memagenta.p
0.03	0.003	0.011



Data Analysis – Untargeted Approach – WGCNA - Conclusion

- We applied WGCNA to a single batch of 300 individuals with about 4,000 metabolites.
- WGCNA is able to identify hub metabolites responsive associated with SCD complications such as leg ulcers.
- We hope to combine all of ~700 samples together to run this analysis and further characterize changes in SCD metabolome that lead to complications.

Conclusions & Perspectives

- Targeted approach looks promising
 - Found that bile acids play a role in the disease progression for SCD complications.
 - Identified significant genomic variants in SCD cohort to were previously reported in healthy Europeans. We also found some variants that seem new.
 - Waiting to collaborators to share their association results in order to combine it with our data.
 - Next step Mendelian Randomization.
- Untargeted approach with WGCNA provides hypothesis generating results
 - The analytic technique clustered metabolites into groups which were specifically related to complications.
 - We need to figure out ways to put collate all of samples and run the analysis.
 - Relate all these modules to SCD complications and pathways (KEGG).

Acknowledgements

- Gabriel Boucher
- Simon Lalonde



Questions ?

SCD Metabolomics Literature

Detrimental effects of adenosine signaling in sickle cell disease. Zhang Y, et al. Nature Med. 2011 Jan;17(1):79-86.

Elevated sphingosine-1-phosphate promotes sickling and sickle cell disease progression. Zhang Y, et al. JCI. 124, 2750-2761, (2014).

Pathophysiology of sickle cell disease is mirrored by the red blood cell metabolome. Darghouth, D. et al. Blood 117, E57-66, (2011).

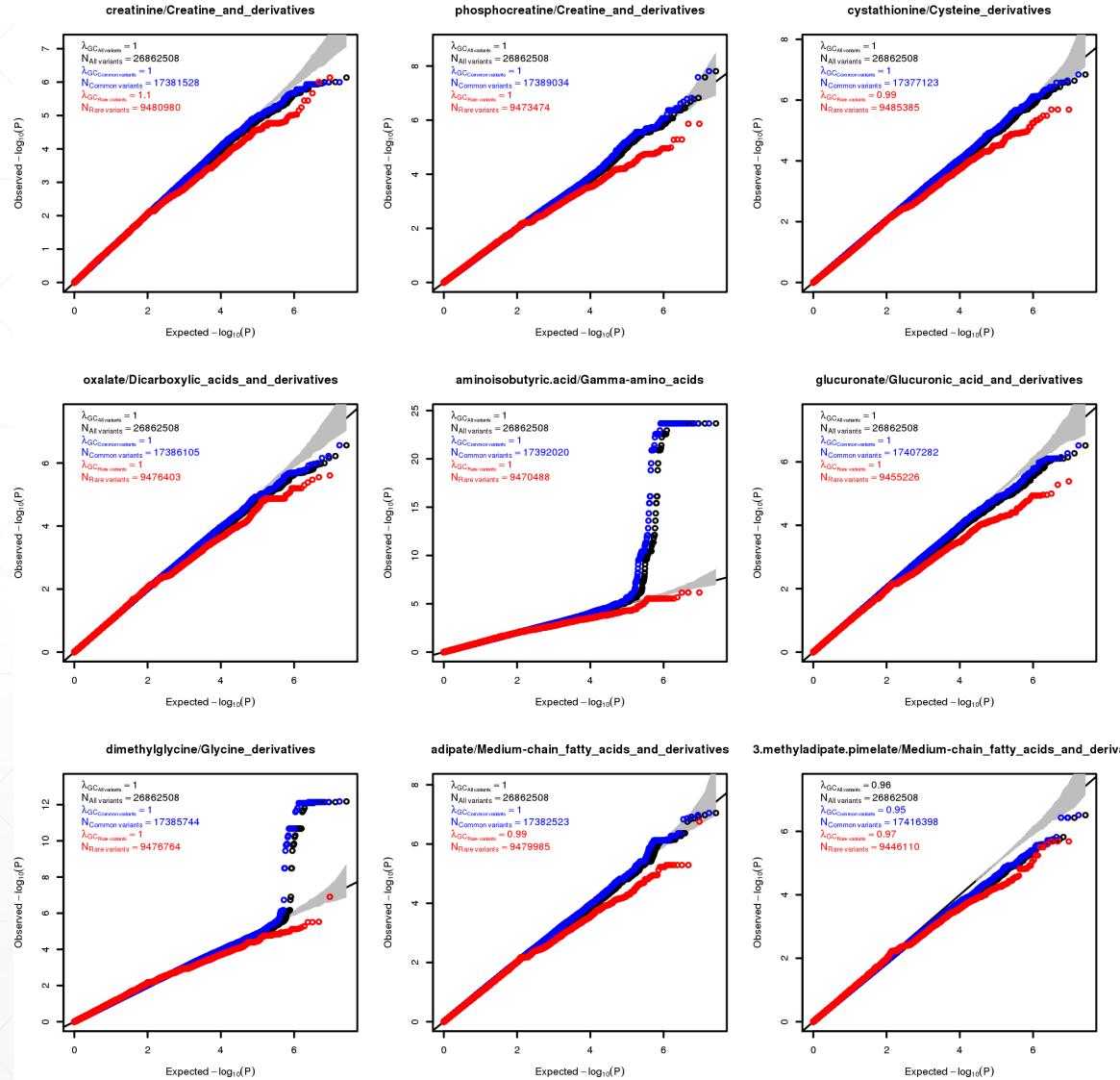
Structural and Functional Insight of Sphingosine 1-Phosphate-Mediated Pathogenic Metabolic Reprogramming in Sickle Cell Disease. Sun K, D'Alessandro A, Ahmed MH, et al. Sci Rep. 2017;7(1):15281.

Appendix

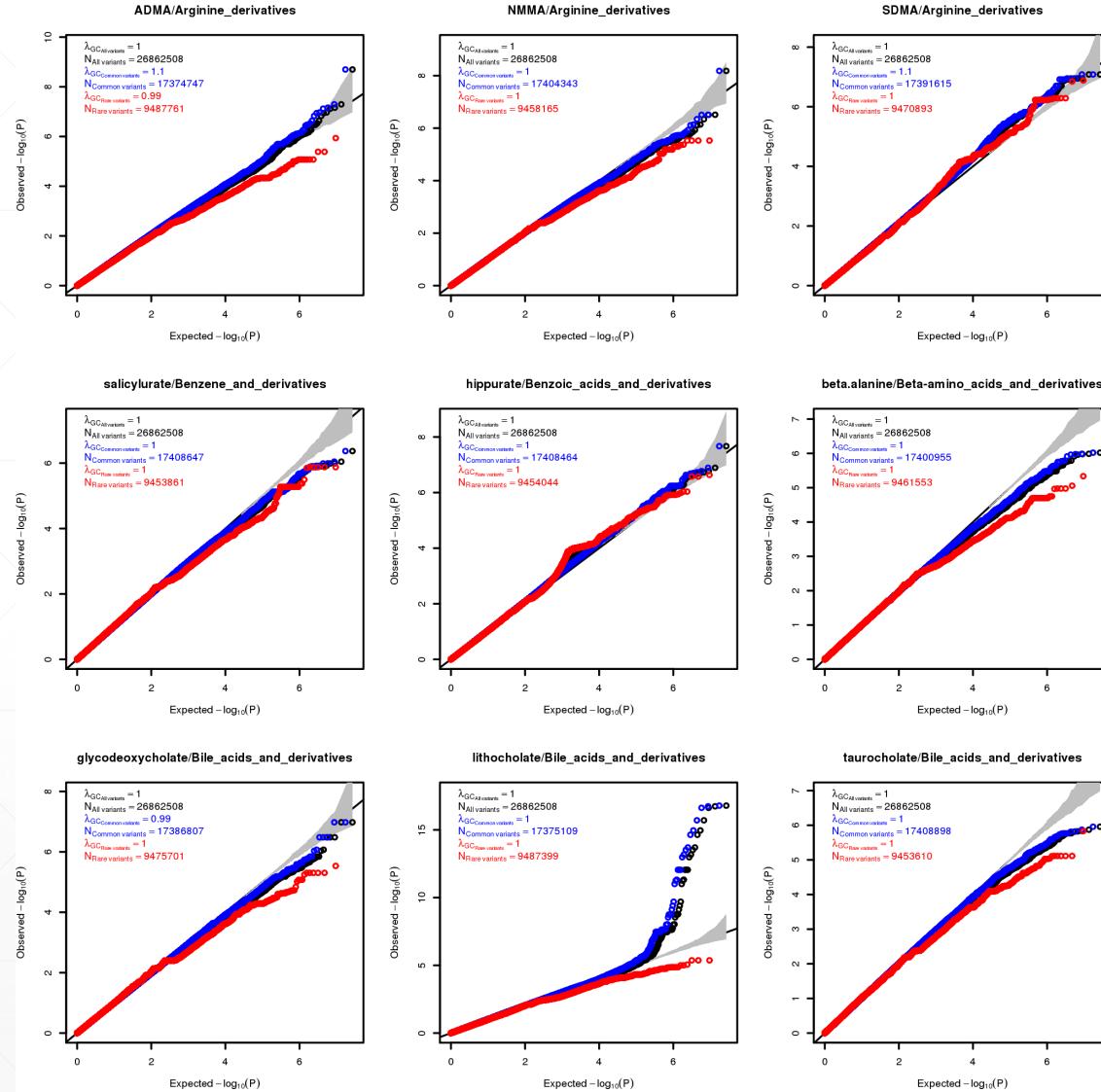
GWAS in Healthy Europeans – Shin. et al. 2014

Sample Collection	Key	TwinsUK	KORA F4
Cohort descriptives	Country of origin	UK	Germany
	N by gender (M/F)	6,056 (433/5,623)	1,768 (858/910)
	Age (years, mean (SD))	53.4 (14.0)	60.8 (8.8)
	BMI (kg/m ² , mean (SD))	26.1 (4.9)	28.2 (4.8)
Metabolite measurements	Sample type (serum)	1,052	1,768
	Sample type (plasma)	5,004	
	N metabolites (known/unknown)	296/207	305 /212

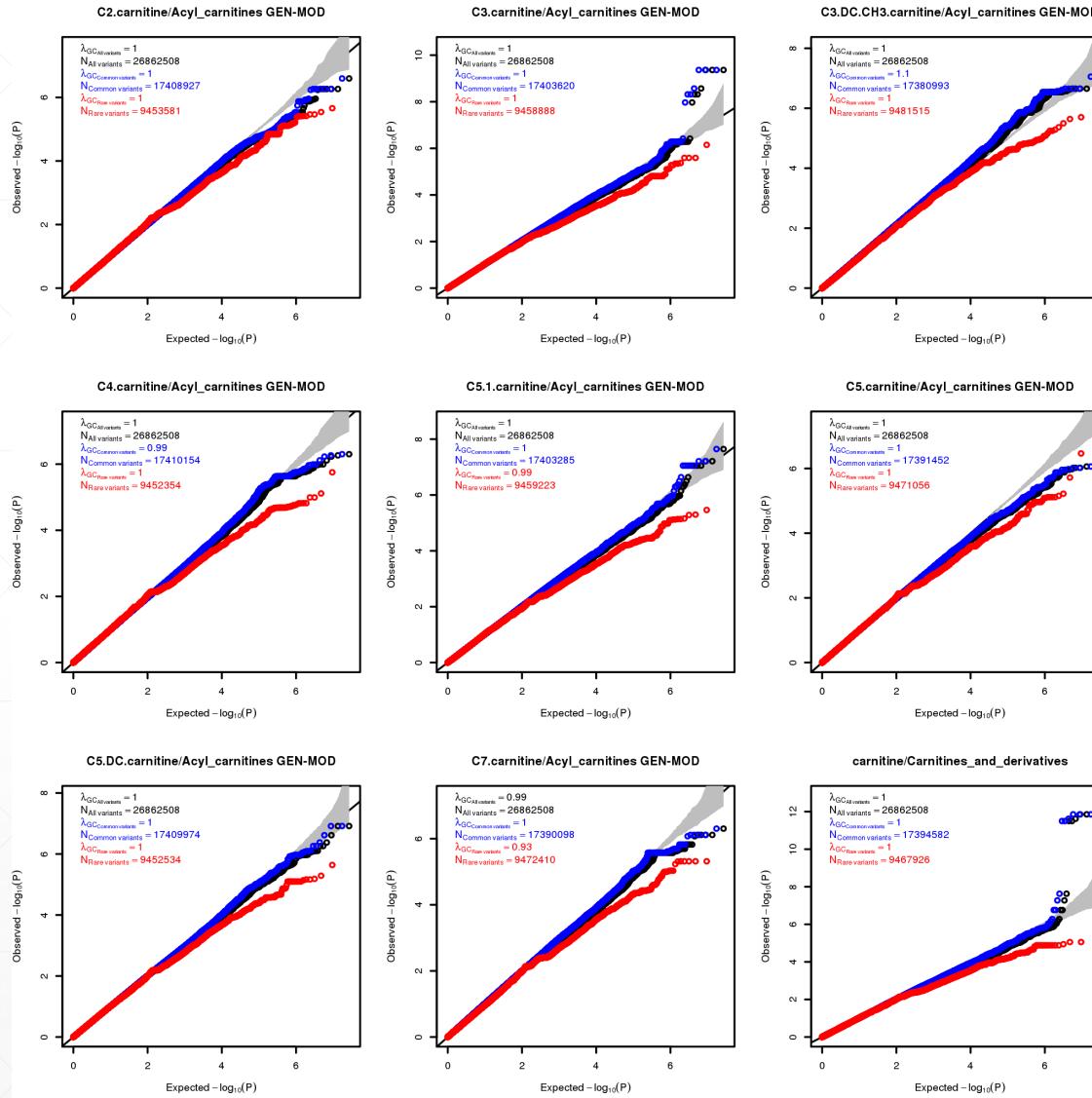
GWAS – QQ-plots



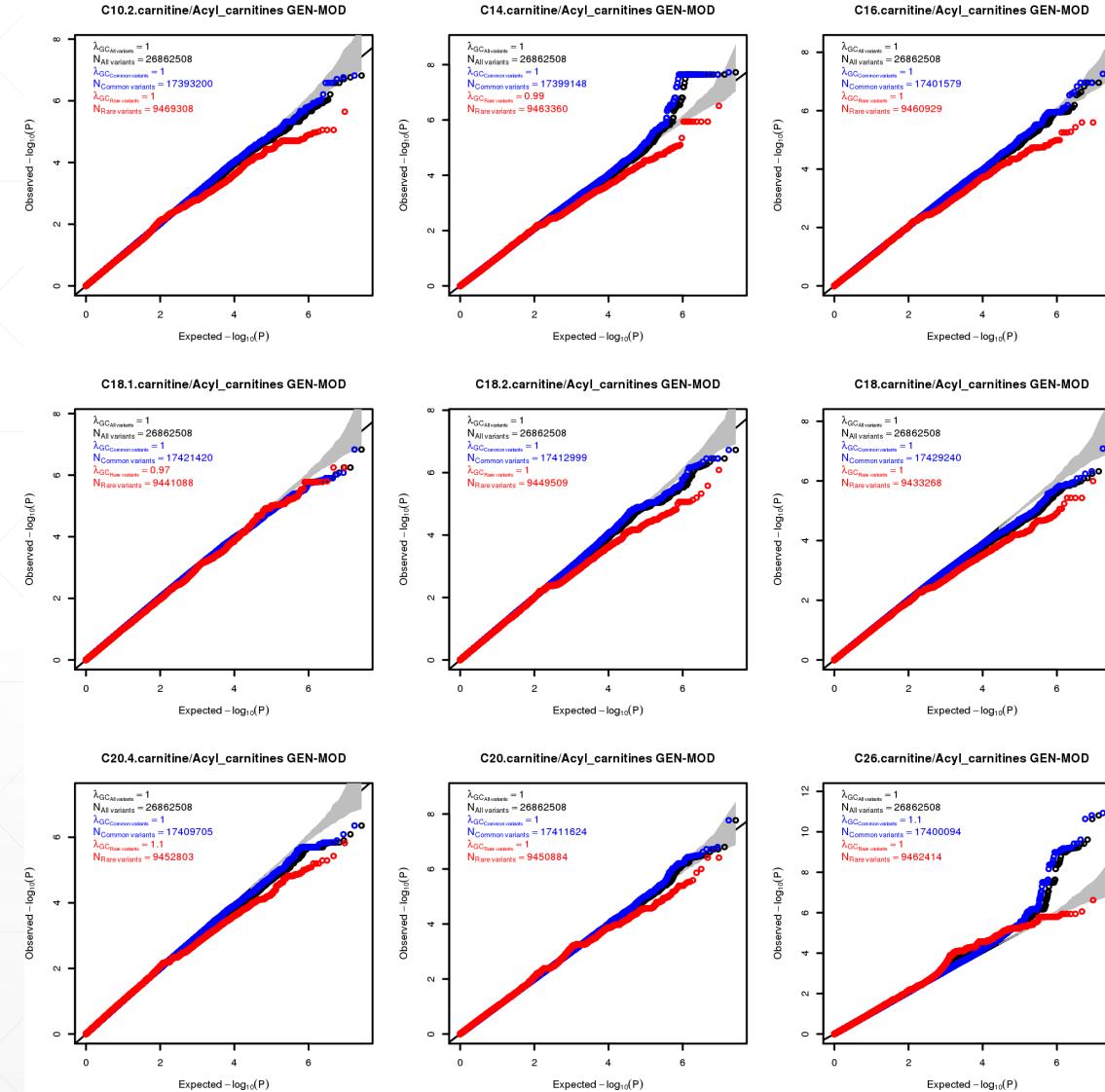
GWAS – QQ-plots



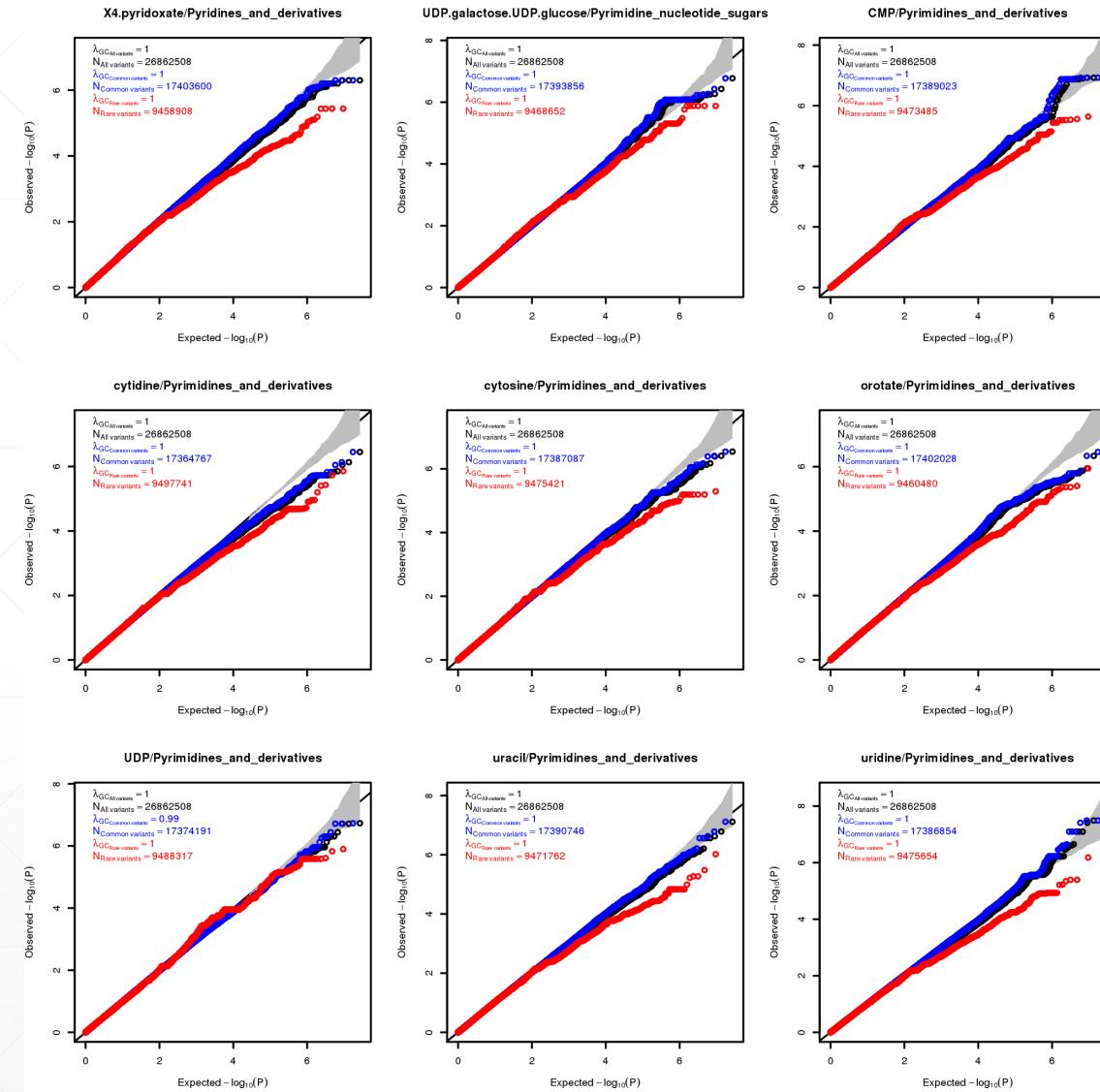
GWAS – QQ-plots



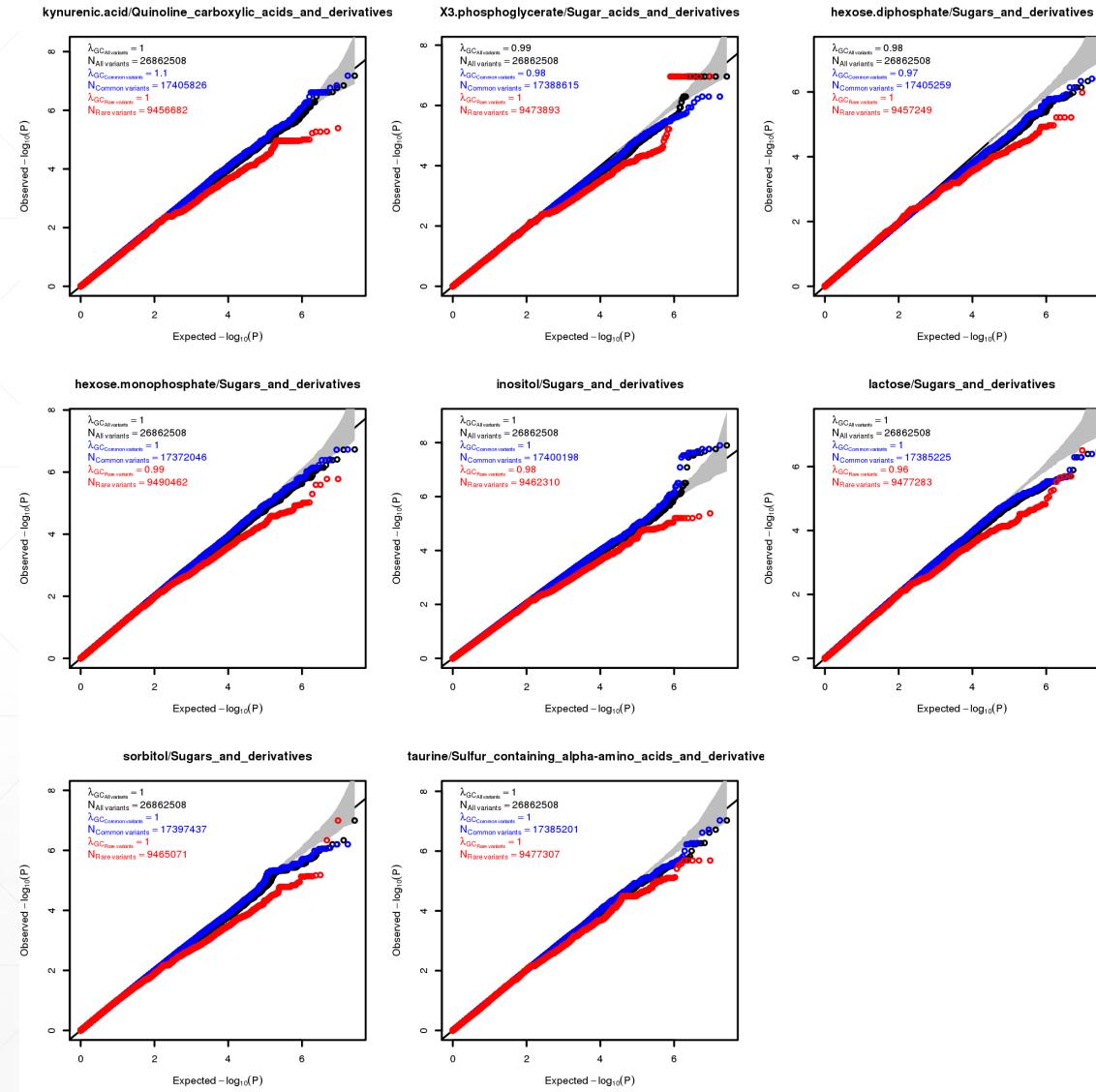
GWAS – QQ-plots



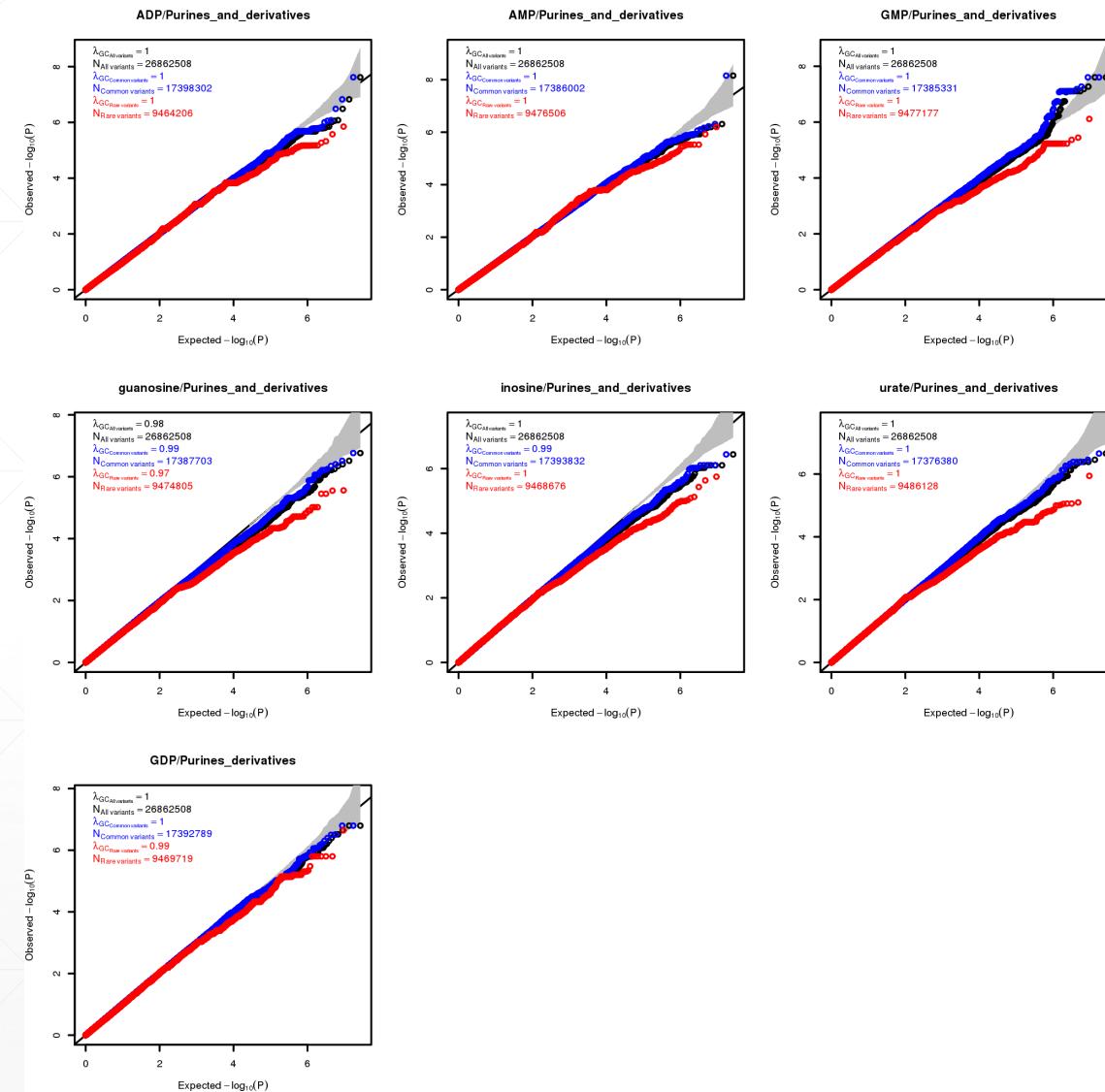
GWAS – QQ-plots



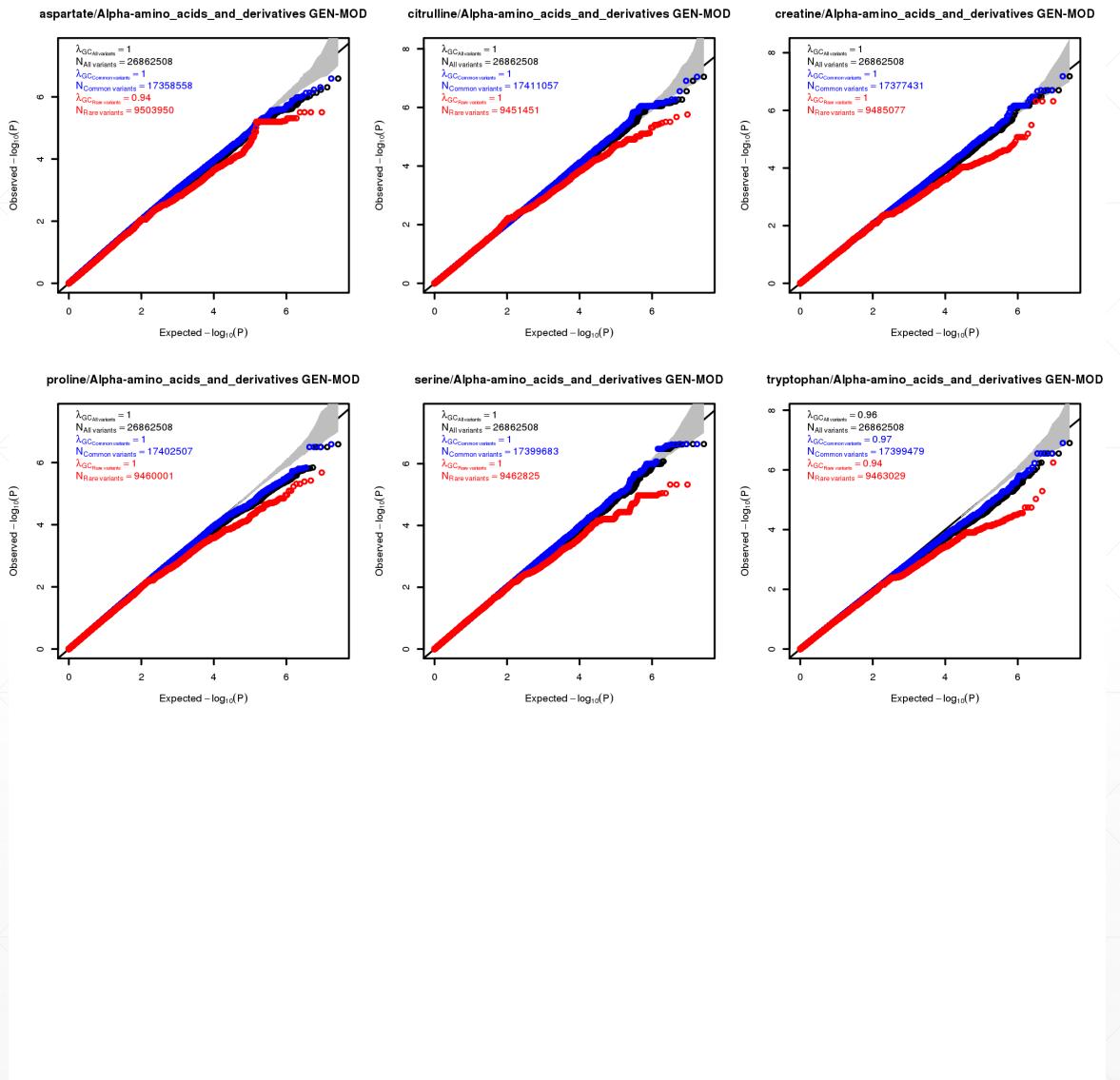
GWAS – QQ-plots



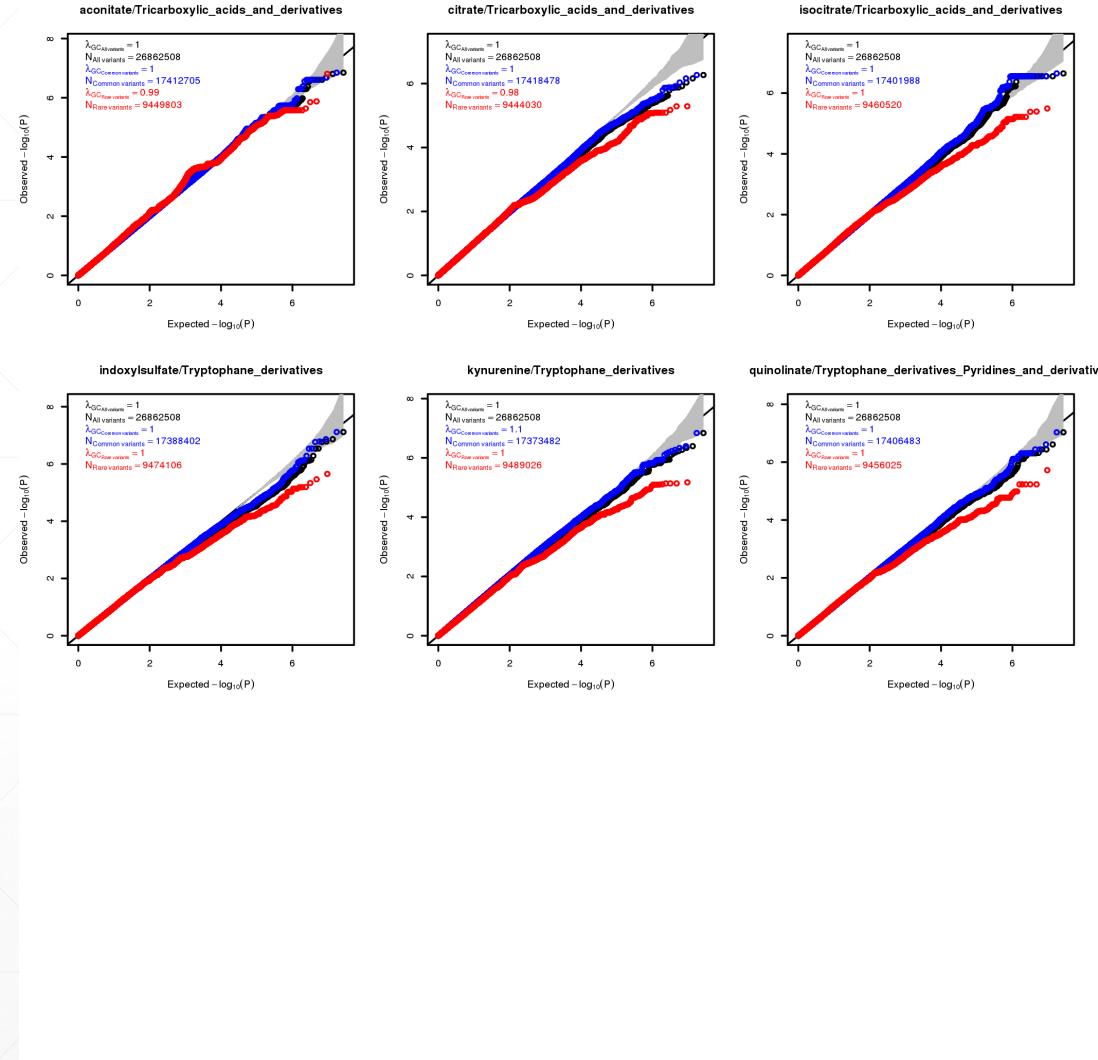
GWAS – QQ-plots



GWAS – QQ-plots



GWAS – QQ-plots



GWAS – QQ-plots

