

HelmholtzZentrum münchen

Deutsches Forschungszentrum für Gesundheit und Umwelt

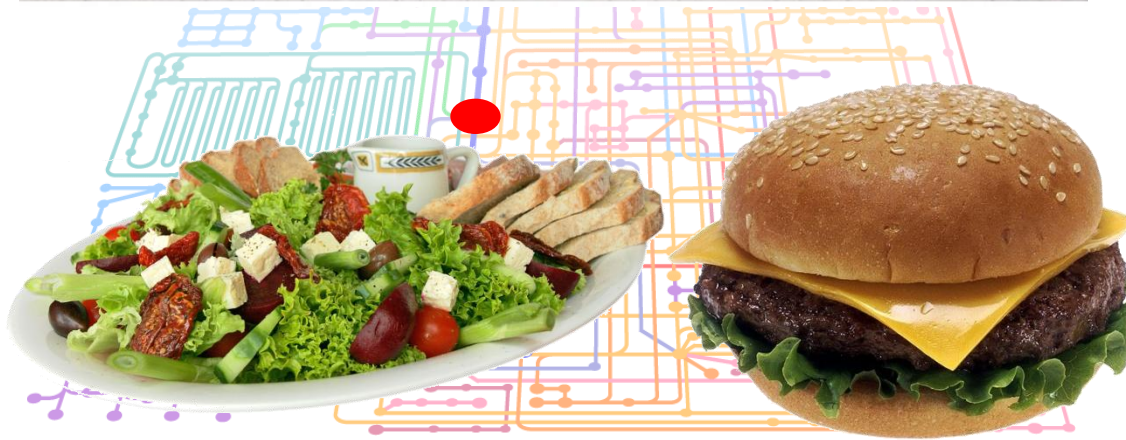
Analysis of metabolomic data in the context of genotype/phenotype information

A Genome-Wide Association Study In Human Urine

Johannes Raffler

Helmholtz Zentrum München
Institute of Bioinformatics and Systems Biology (IBIS)

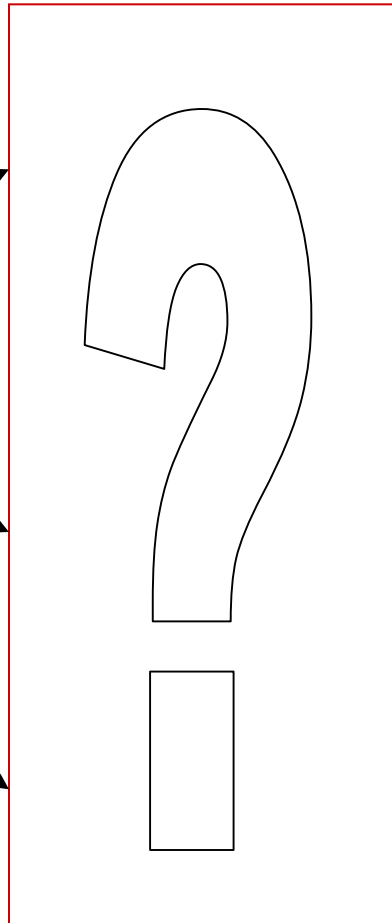
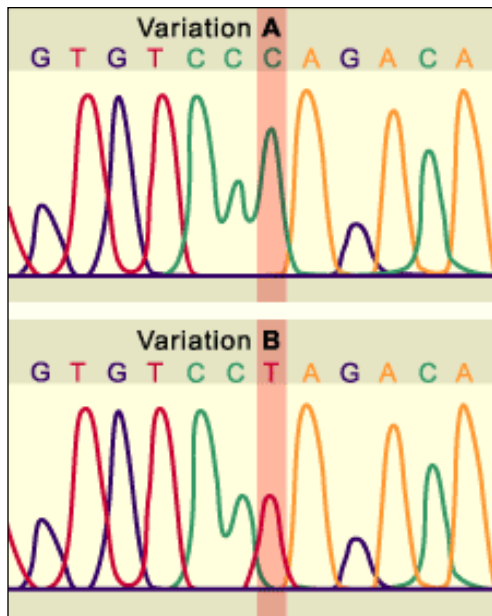
What makes us different?



From genotype to phenotype

Phenotype

Genotype (SNPs)



disease related phenotypes



cardio vascular diseases



obesity



allergy



lifestyle ...

GWA of metabolic traits in urine

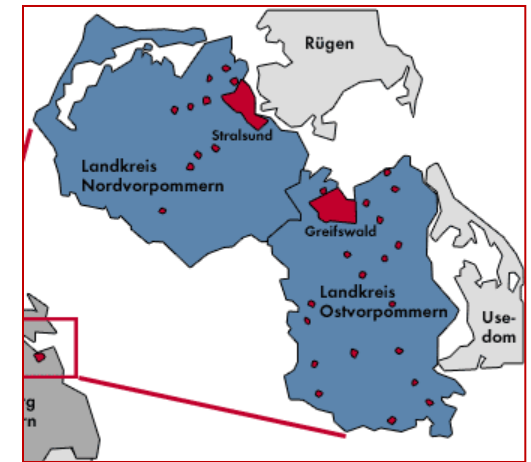
The SHIP study

Study design:

- Study of **H**ealth in **P**omerania
- conducted by the University of Greifswald
- out of > 200,000 inhabitants in that area, 4,308 participants were sampled (both genders)
- sampling stages:
SHIP-0: 1997 - 2001
SHIP-1: 2002 - 2006

Data provided to us:

1. Genomic data (SNPs)
2. NMR spectra of urine samples



Processing the data



Genomic data:

- Original data: 909,508 SNPs per person determined
- After filtering: 645,249 SNPs (MAF > 5%, call rate > 95%, p(HWE) > 0.001)
- Result: for each SNP it is determined whether the person is a *major* homozygote (0 minor alleles), heterozygote (1) or *minor* homozygote (2)

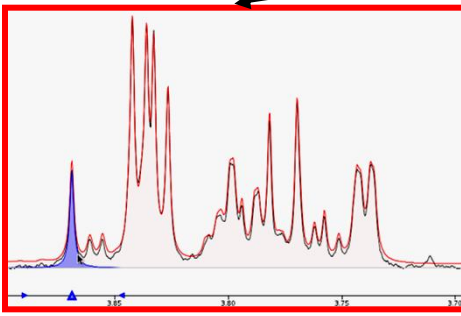
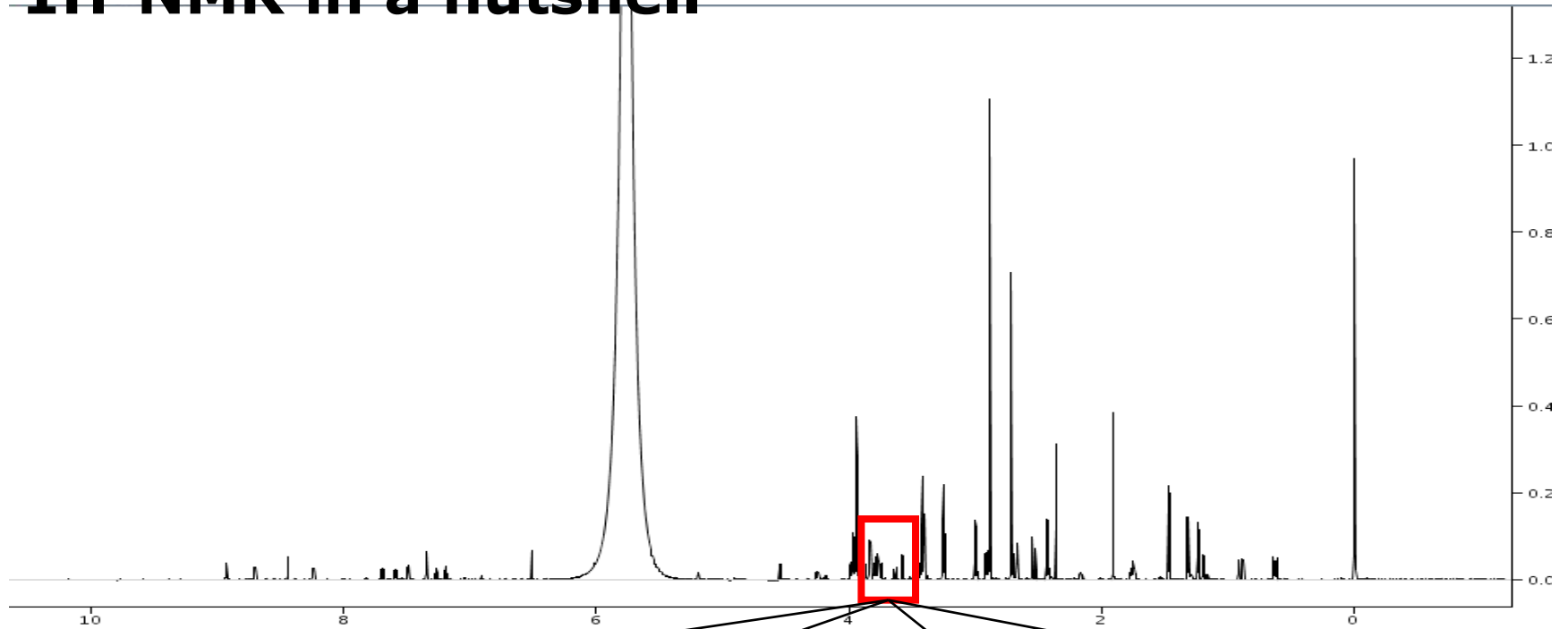
Metabolomic data:

- Original data: raw NMR spectra
- Goal: quantification of the metabolites present in the urine samples

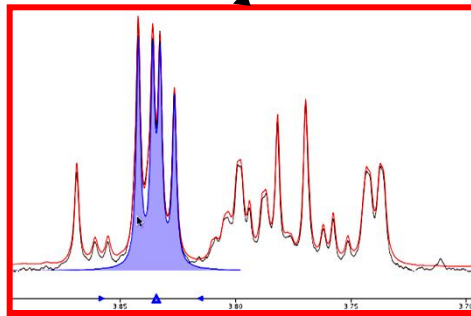
but how?



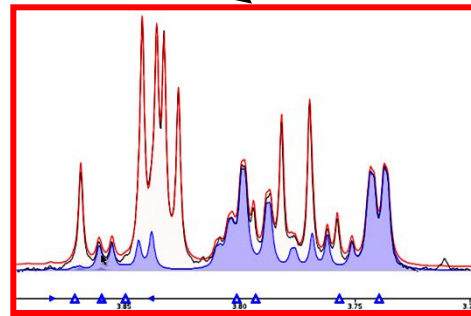
1H-NMR in a nutshell



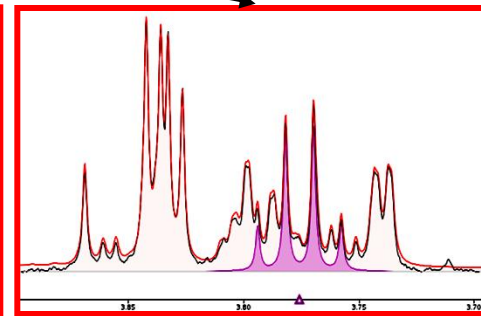
4-Hydroxy-3-methoxymandelate



Serine



Fucose

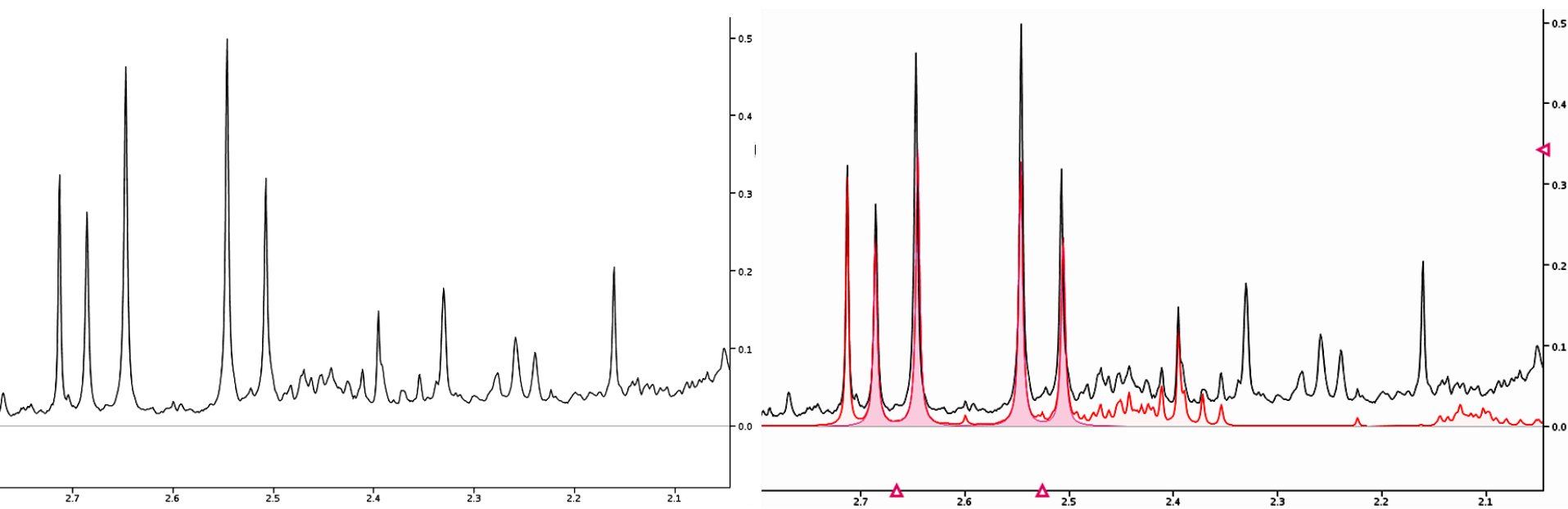


Alanine

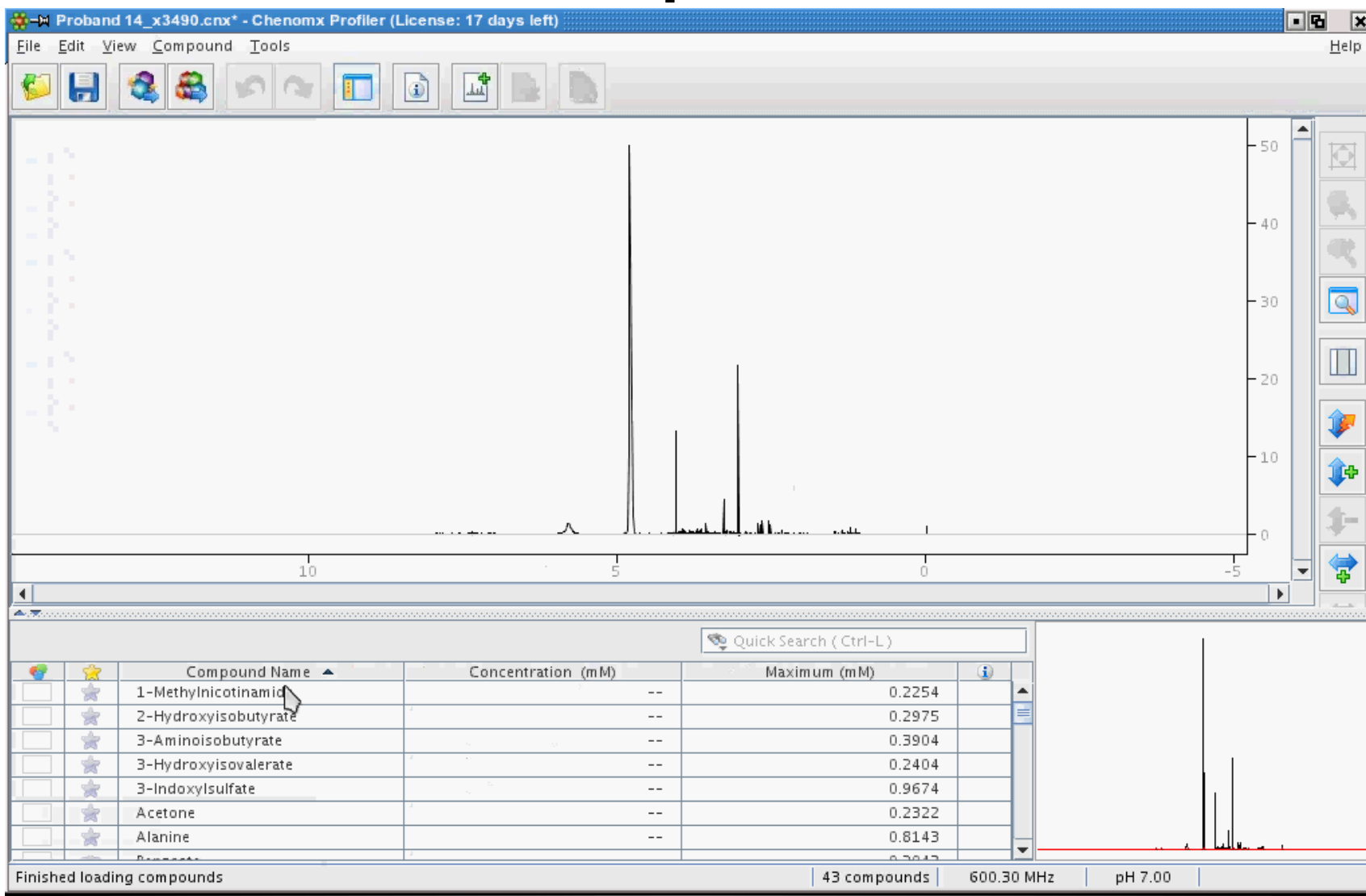
How to annotate NMR spectra

One approach:

NMR profiling using “fingerprints” of different compounds



How to annotate NMR spectra



Performing the GWA

After the annotation:

- from the SHIP-0 data set (partially fee-for-service)
- 63 metabolites were quantified

hetero-
zygotes

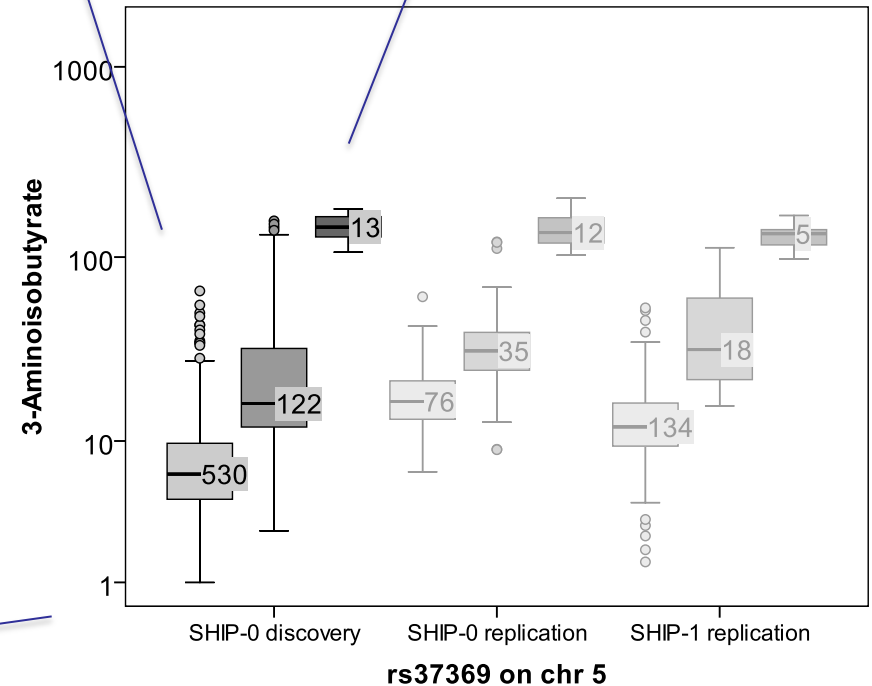
were annotated

minor
homozygotes

The next step:

- for each SNP:
 - look for associations between
 - minor allele count and
 - metabolites concentrations
- this was automated using PLINK

major
homozygotes



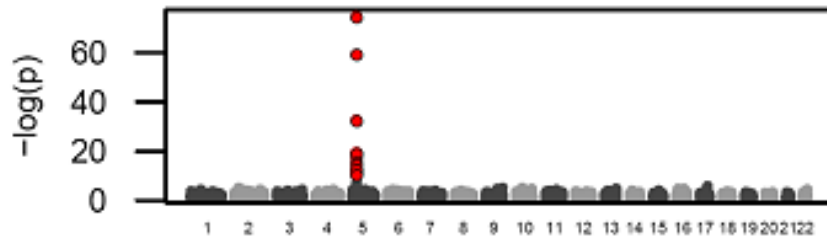
GWA results

- testing **1720 metabolic traits** (59 metabolites and the „all-against-all“ ratios) to 645,249 SNPs:
 - 1,319 associations to 161 SNPs significant
 - 161 SNPs represent 61 different genetic loci (within 1 Mb)
- correction for multiple testing of 1720 traits ($p < 4,51e-11 = 0,05 / 645,249 / 1720$), **five associations remain significant**

rs-number	metabolic trait	SHIP-0 discovery			SHIP-0 replication			SHIP-1 replication			joint study		
		N	p-value	r ²	N	p-value	r ²	N	p-value	r ²	N	p-value	r ²
rs37369	3-Aminoisobutyrate	665	3.2x10 ⁻⁷⁵	0.399	123	8.3x10 ⁻²⁷	0.618	157	1.5x10 ⁻²³	0.479	945	2.5x10 ⁻¹¹⁵	0.426
rs4921914	Formate / Succinate	854	5.1x10 ⁻¹⁶	0.074	126	1.7x10 ⁻⁴	0.109	170	0.019	0.032	1150	2.0x10 ⁻¹⁹	0.069
rs830124	2-Hydroxyisobutyrate	859	1.6x10 ⁻¹⁵	0.072	126	0.081	0.025	169	1.4x10 ⁻⁴	0.084	1154	1.1x10 ⁻¹⁷	0.062
rs8101881	Lysine / Valine	605	1.8x10 ⁻¹⁴	0.093	121	0.993	0.000	145	5.4x10 ⁻⁶	0.136	871	6.8x10 ⁻¹⁸	0.082
rs17279437	Alanine / N,N-Dimethylglycine	831	2.3x10 ⁻¹³	0.063	126	0.183	0.014	167	2.8x10 ⁻³	0.053	1124	3.4x10 ⁻¹⁴	0.050

GWA using PLINK

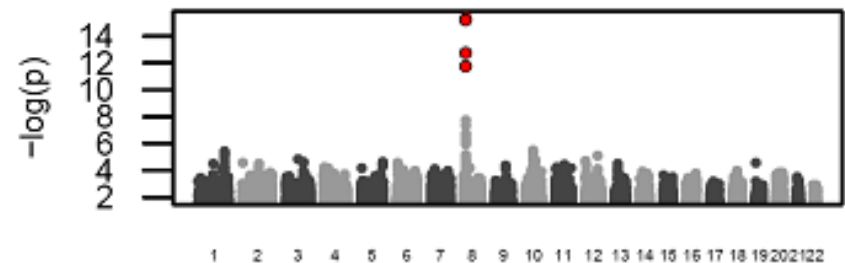
3-Aminoisobutyrate



chromosome position

(a)

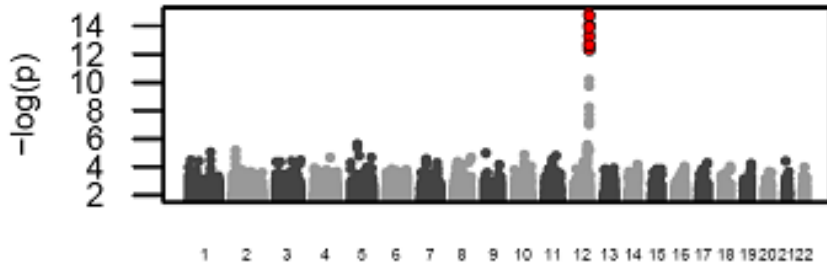
Formate / Succinate



chromosome position

(b)

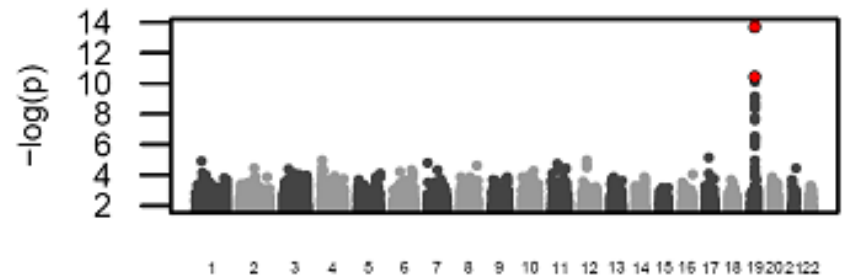
2-Hydroxyisobutyrate



chromosome position

(c)

Lysine / Valine



chromosome position

(d)

Replication

SHIP-0:

- ~ 130 male individuals:
 - urine sampling done after 11 a.m.
 - minor allele homozygotes for at least one of the SNPs of interest

SHIP-1:

- ~ 170 male individuals:
 - same persons as in the SHIP-0 dataset
 - minor allele homozygotes for at least one of the SNPs

Replication

rs-number	metabolic trait	SHIP-0 discovery			SHIP-0 replication			SHIP-1 replication			joint study		
		N	p-value	r ²	N	p-value	r ²	N	p-value	r ²	N	p-value	r ²
rs37369	3-Aminoisobutyrate	665	3.2x10 ⁻⁷⁵	0.399	123	8.3x10 ⁻²⁷	0.618	157	1.5x10 ⁻²³	0.479	945	2.5x10 ⁻¹¹⁵	0.426
rs4921914	Formate / Succinate	854	5.1x10 ⁻¹⁶	0.074	126	1.7x10 ⁻⁴	0.109	170	0.019	0.032	1150	2.0x10 ⁻¹⁹	0.069
rs830124	2-Hydroxyisobutyrate	859	1.6x10 ⁻¹⁵	0.072	126	0.081	0.025	169	1.4x10 ⁻⁴	0.084	1154	1.1x10 ⁻¹⁷	0.062
rs8101881	Lysine / Valine	605	1.8x10 ⁻¹⁴	0.093	121	0.993	0.000	145	5.4x10 ⁻⁶	0.136	871	6.8x10 ⁻¹⁸	0.082
rs17279437	Alanine / N,N-Dimethylglycine	831	2.3x10 ⁻¹³	0.063	126	0.183	0.014	167	2.8x10 ⁻³	0.053	1124	3.4x10 ⁻¹⁴	0.050

SHIP-0:
only 2 of 5
associations
could be
replicated

SHIP-1:
4 of 5
associations
could be
replicated

Interpretation

Best hit:

- **SNP rs37369 in AGTXT2**
is the genetic basis for hyper-beta-aminoisobutyric aciduria,
arguably the most common Mendelian metabolic variant in man.

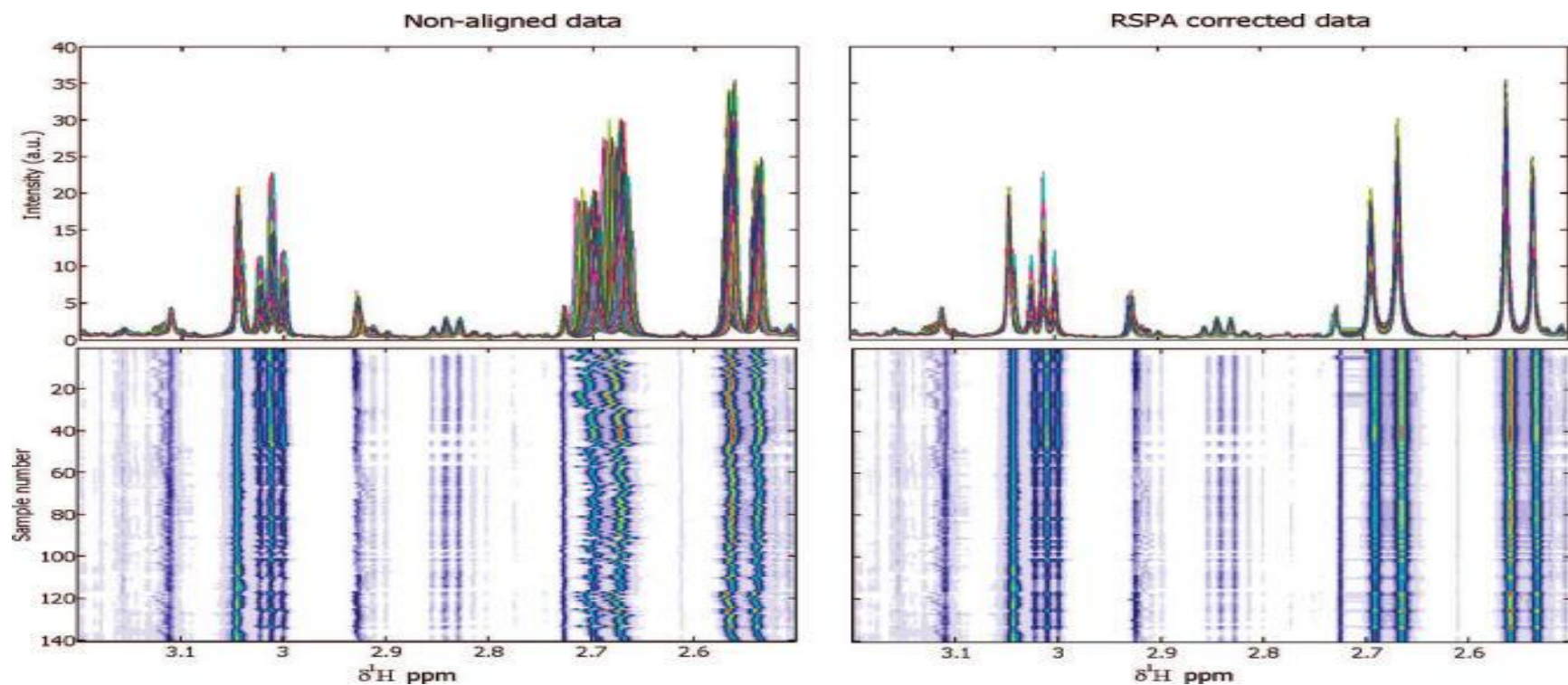
Enzyme / transporter genes known to associate with clinical outcomes:

- SLC7A9 is a risk factor for chronic kidney diseases
- NAT2 confers a genotype dependant response to drug toxicity
- SLC6A20 contributing factor to iminoglycinuria

What's next?

Problem: manual annotation is time-consuming and incomplete

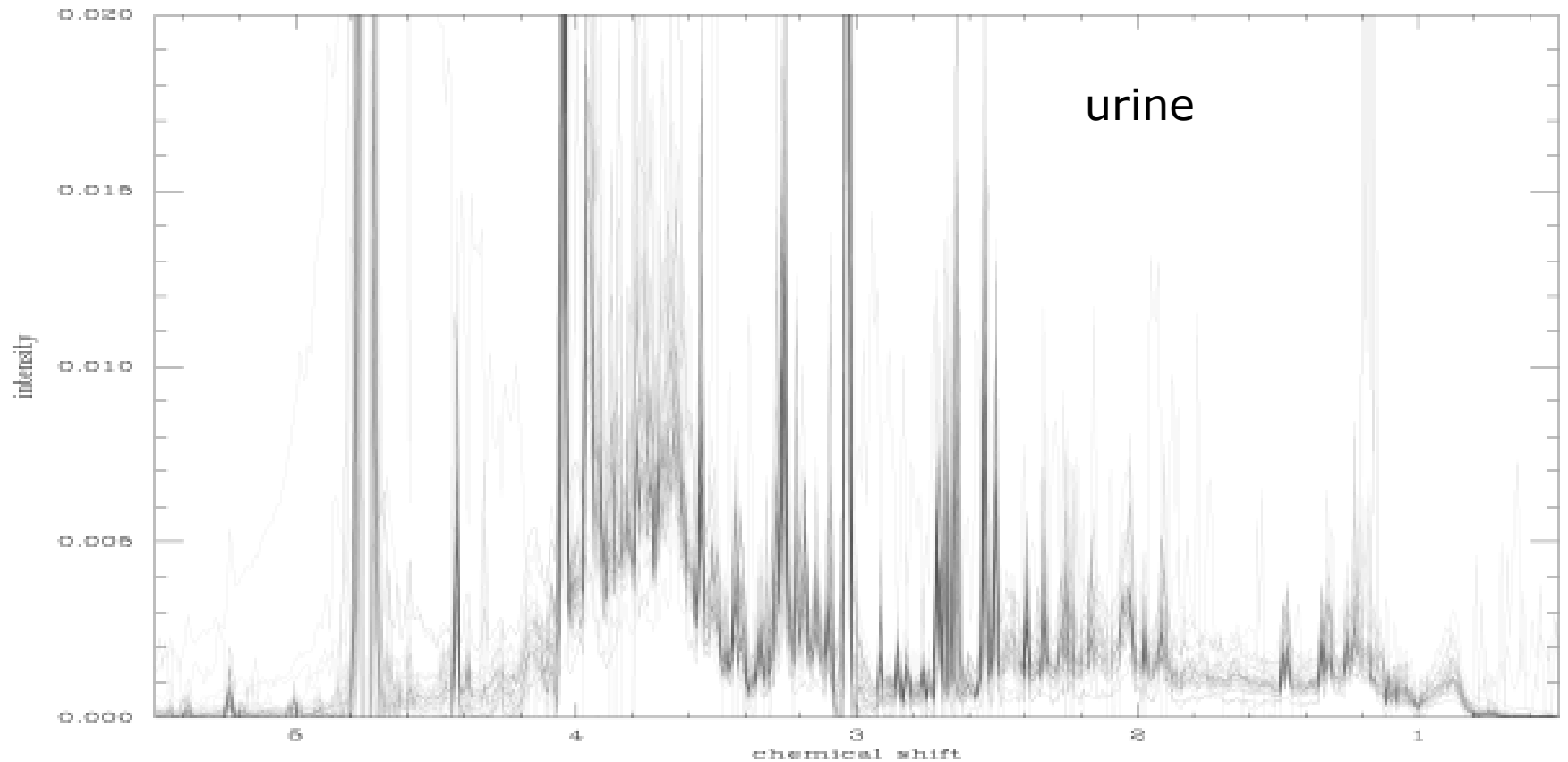
Alternative: do the associations directly at the peak-level?



Nicholson et al., *Anal. Chem.*, 2009, 81 (1), pp 56–66

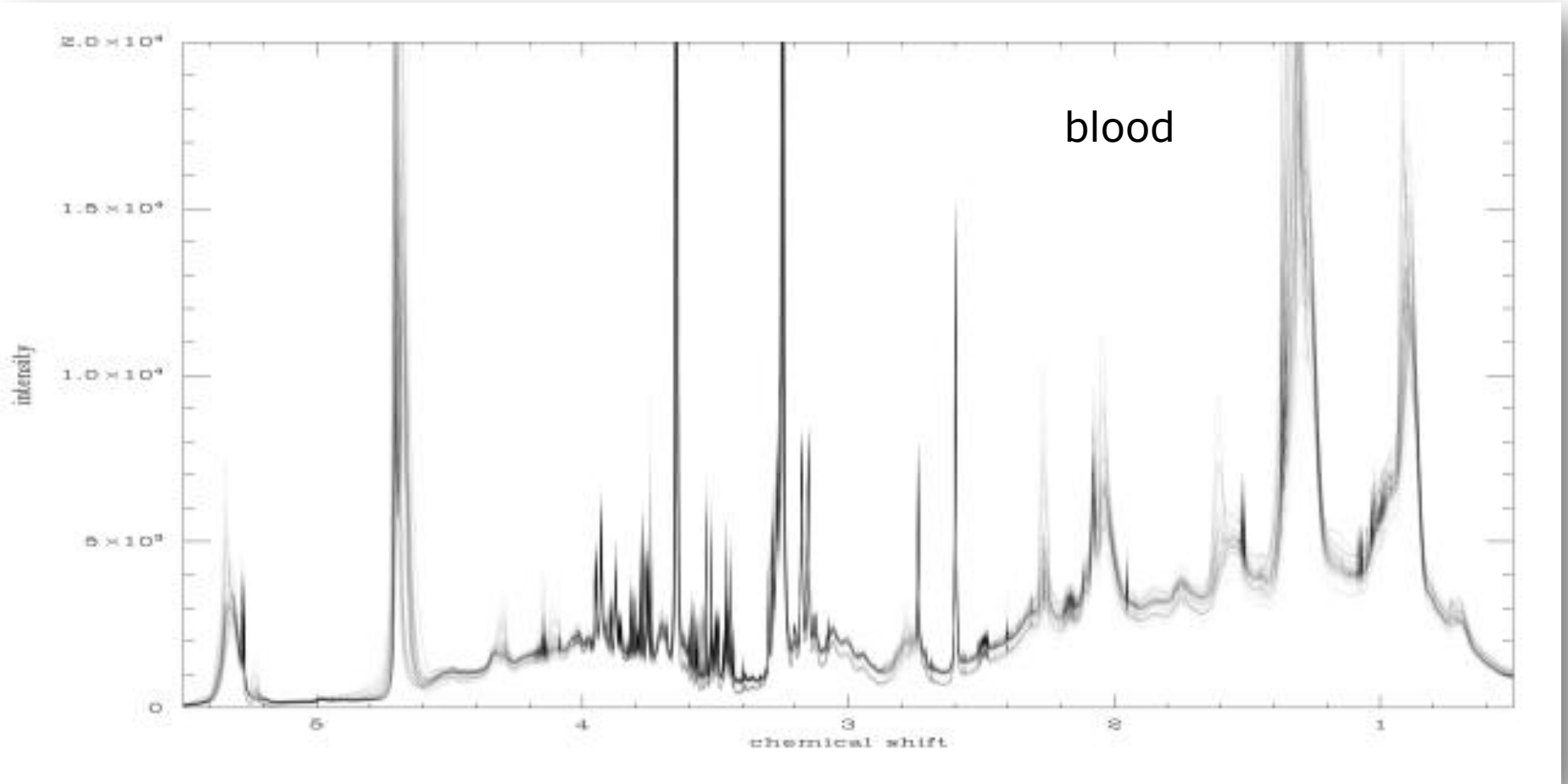
What's next?

Another project: GWA in blood serum (KORA study, 1,800 samples)



What's next?

Another project: GWA in blood serum (KORA study, 1,800 samples)



Thank you!