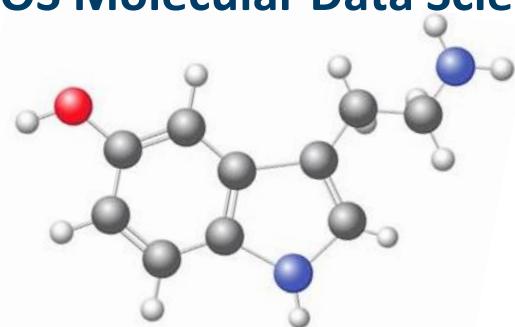


Metabolomics

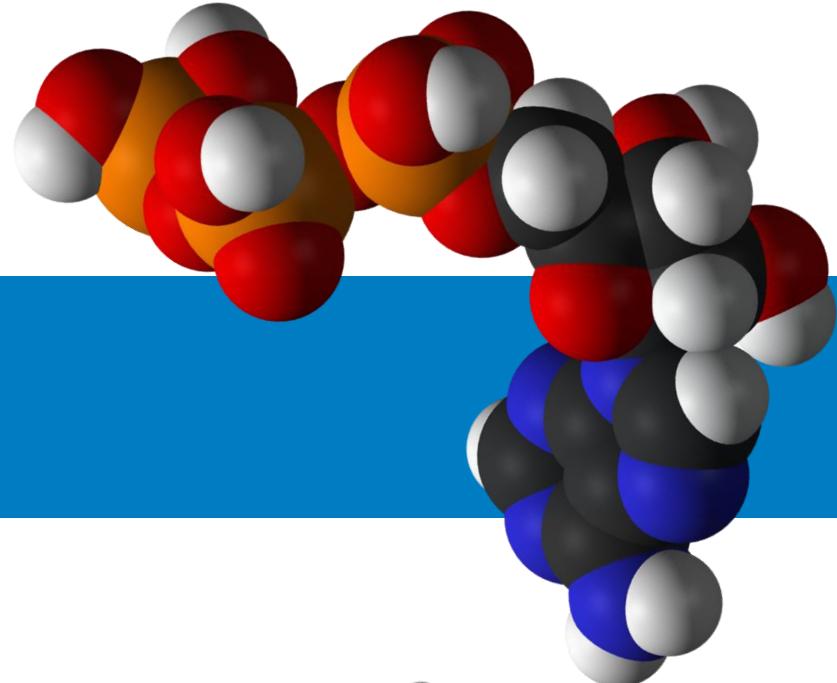
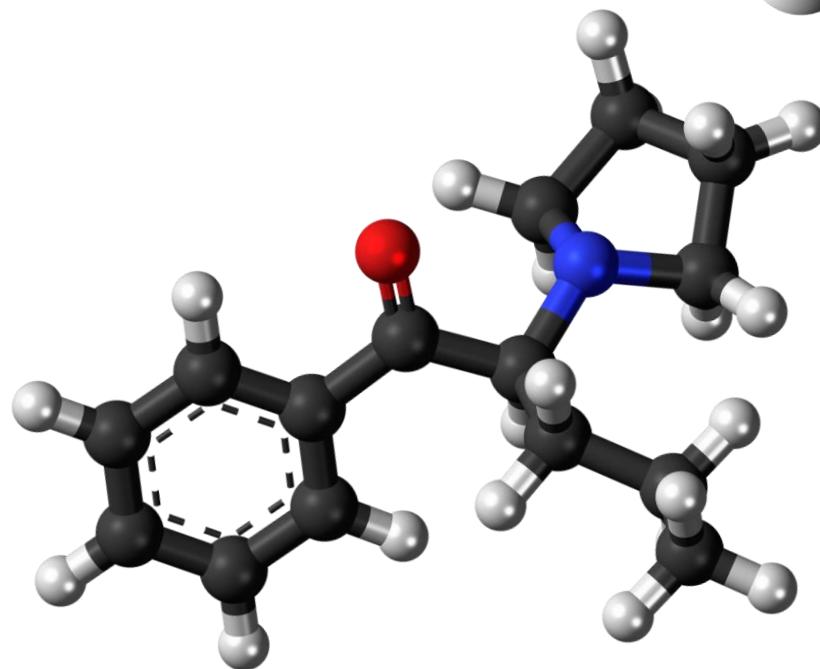
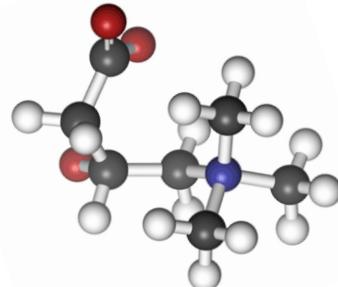
FOS Molecular Data Science 2019



Marian Beekman

Molecular Epidemiology

LUMC



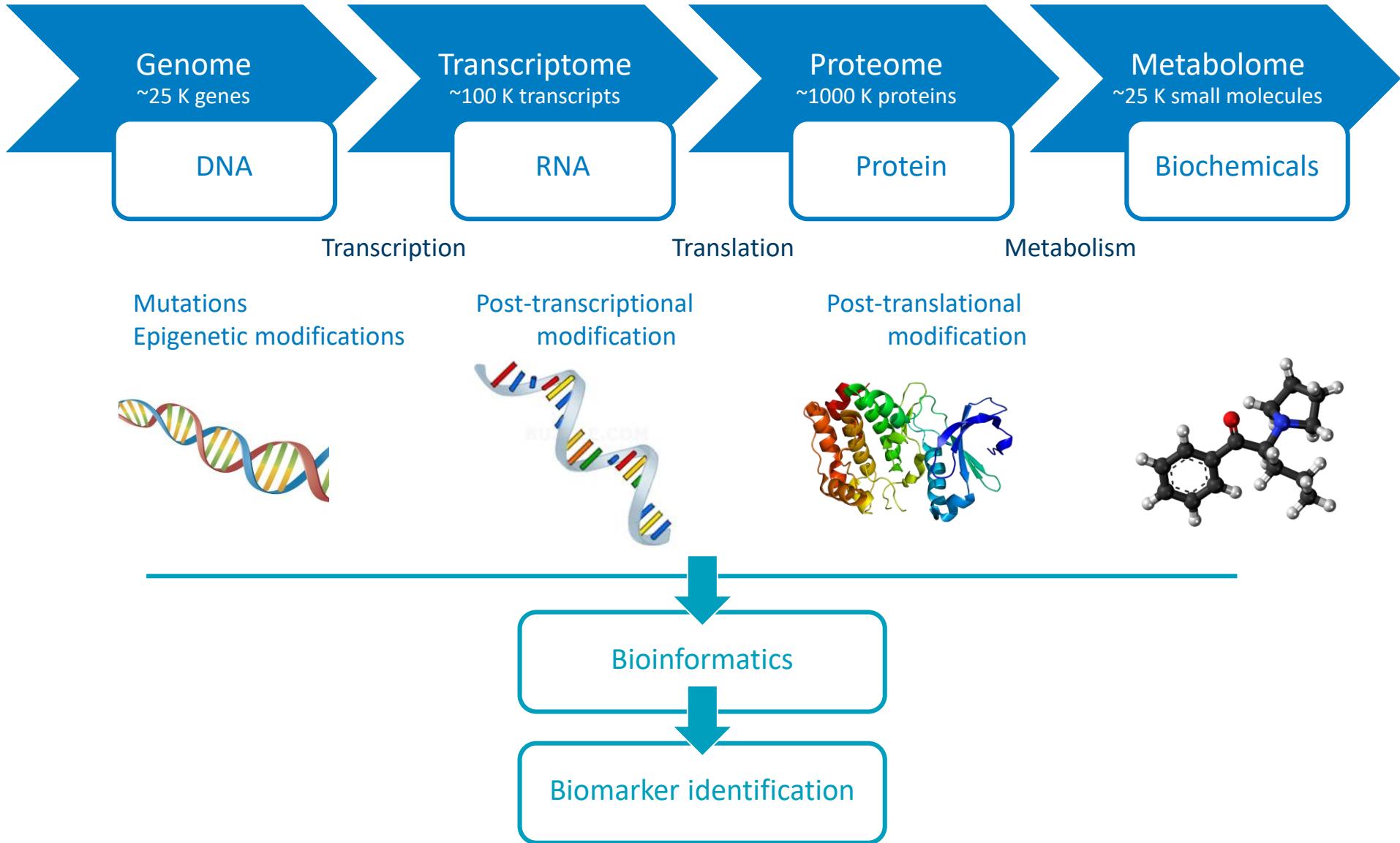
Learning goals

After todays lectures and practical you

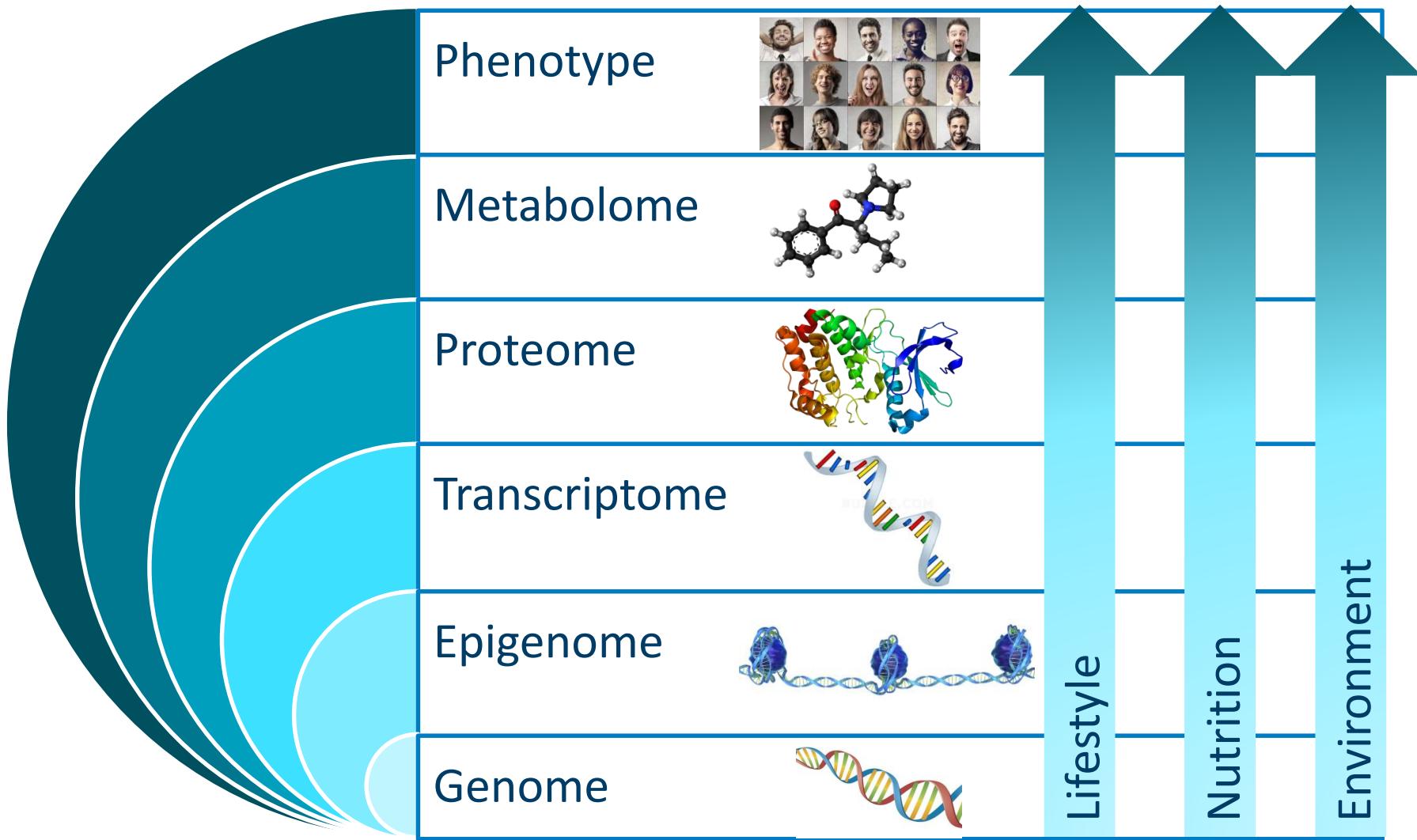
- 1) Can explain the different levels of omics data that are used for disease prediction
- 2) Can explain the strength of the use of metabolomics in disease prediction
- 3) Can explain the weaknesses of the use of metabolomics in disease prediction
- 4) Understand the workflow of -omics data analysis
- 5) Can explain relevant steps in quality control of metabolomics
- 6) Know which databases may contribute to the interpretation of results



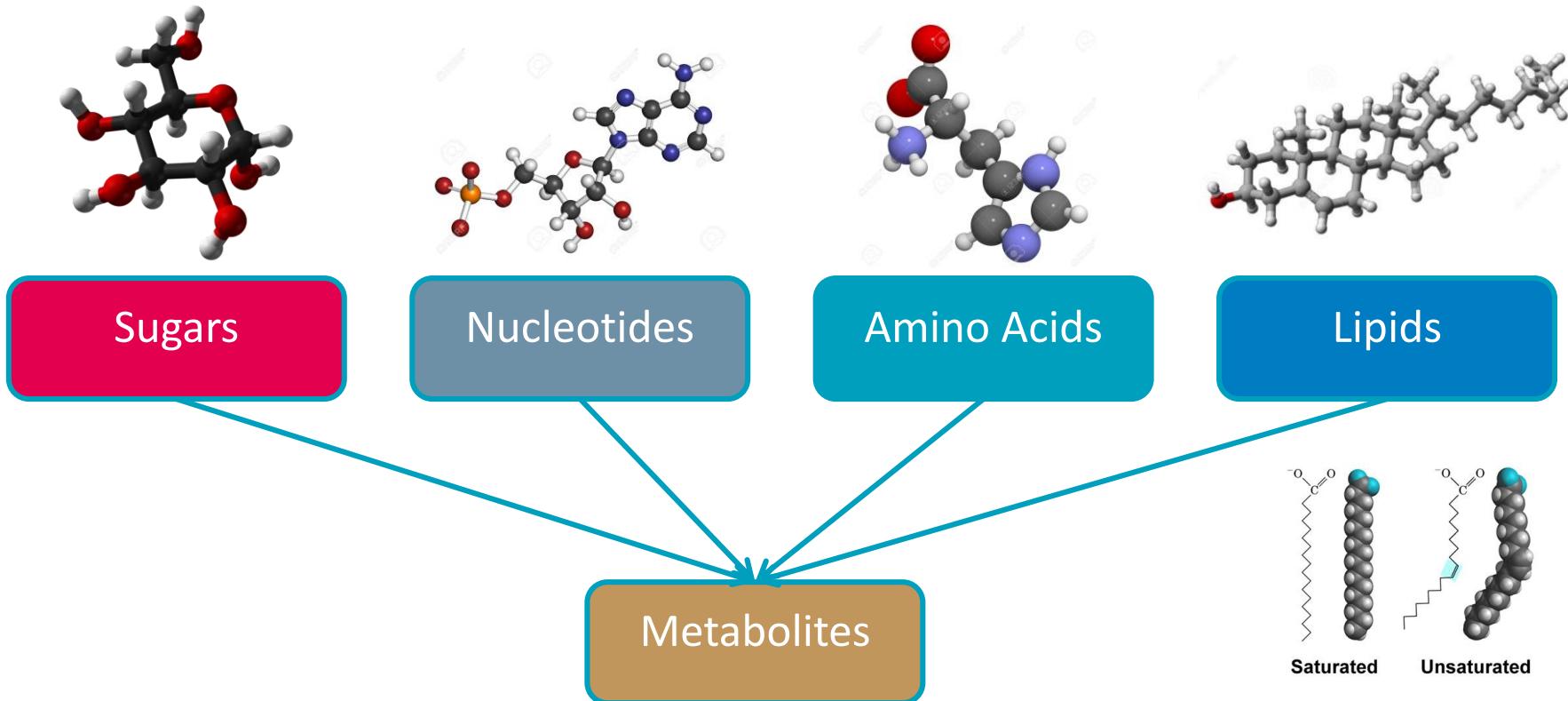
The central dogma of biology



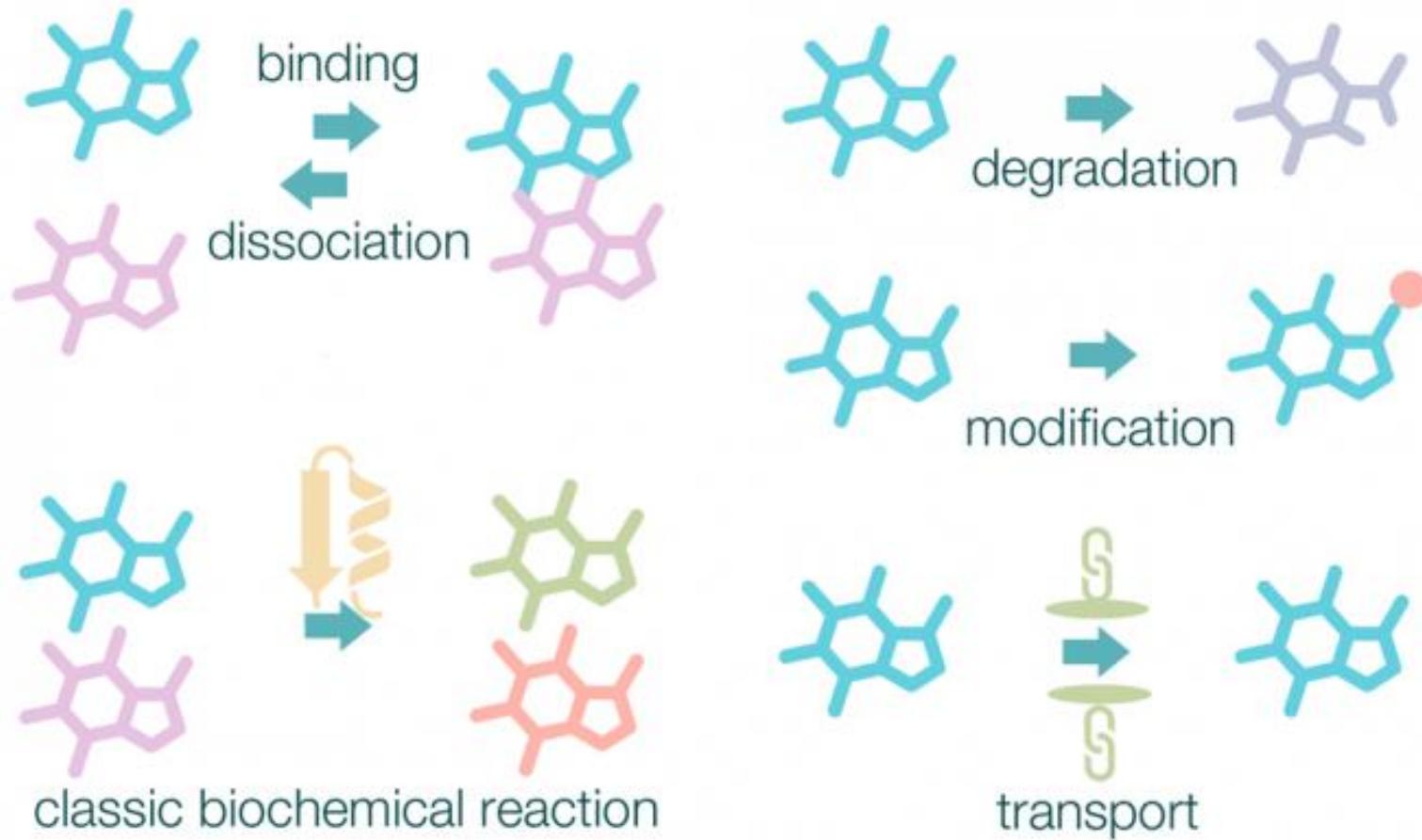
Different -omics levels



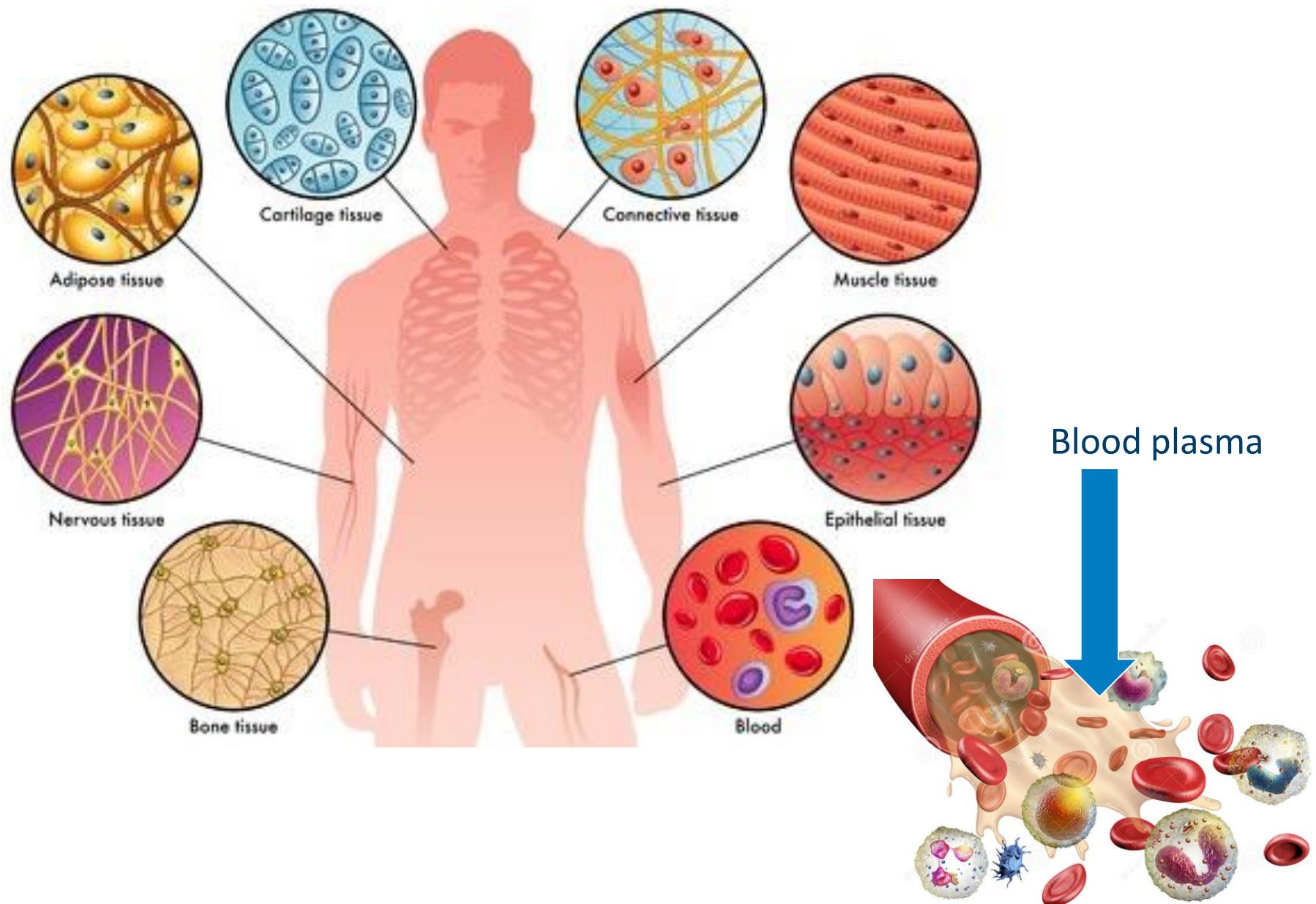
Metabolome



Snapshot in time...



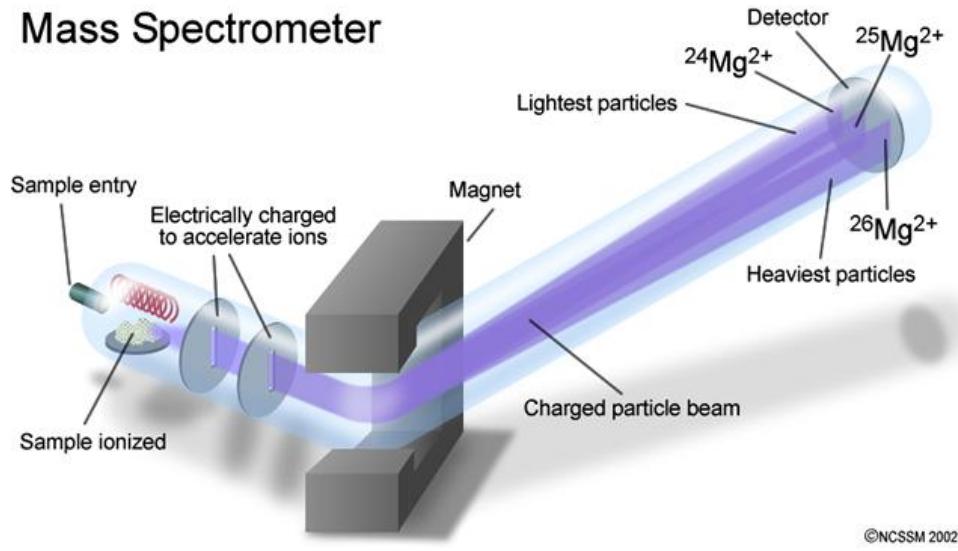
Snapshot in tissues...



Methods to measure metabolomics

- Mass Spectrometry (separation on mass)
- Nuclear Magnetic Resonance (separation on the capacities of absorption and re-emission of electromagnetic radiation)

Mass Spectrometer

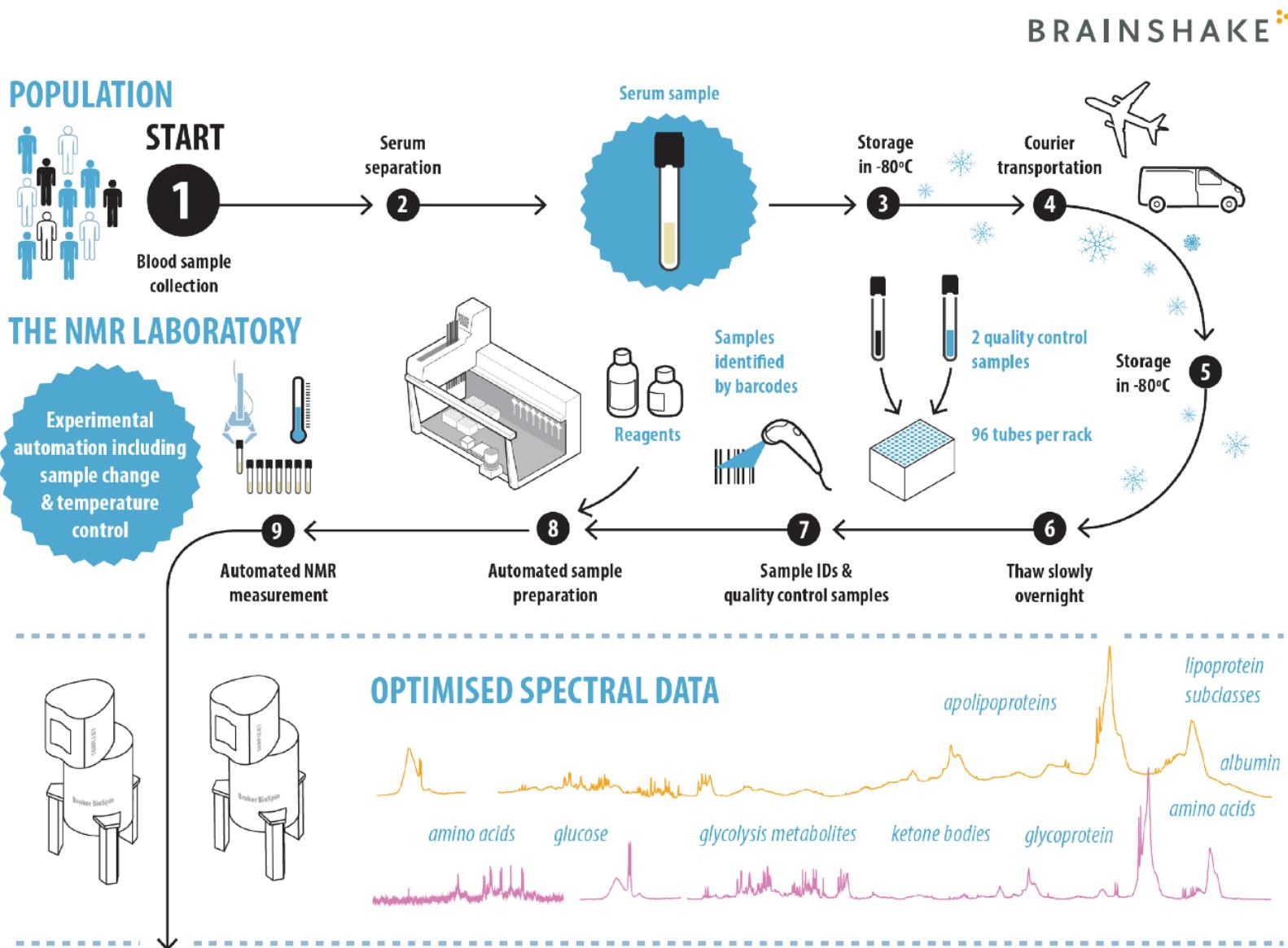




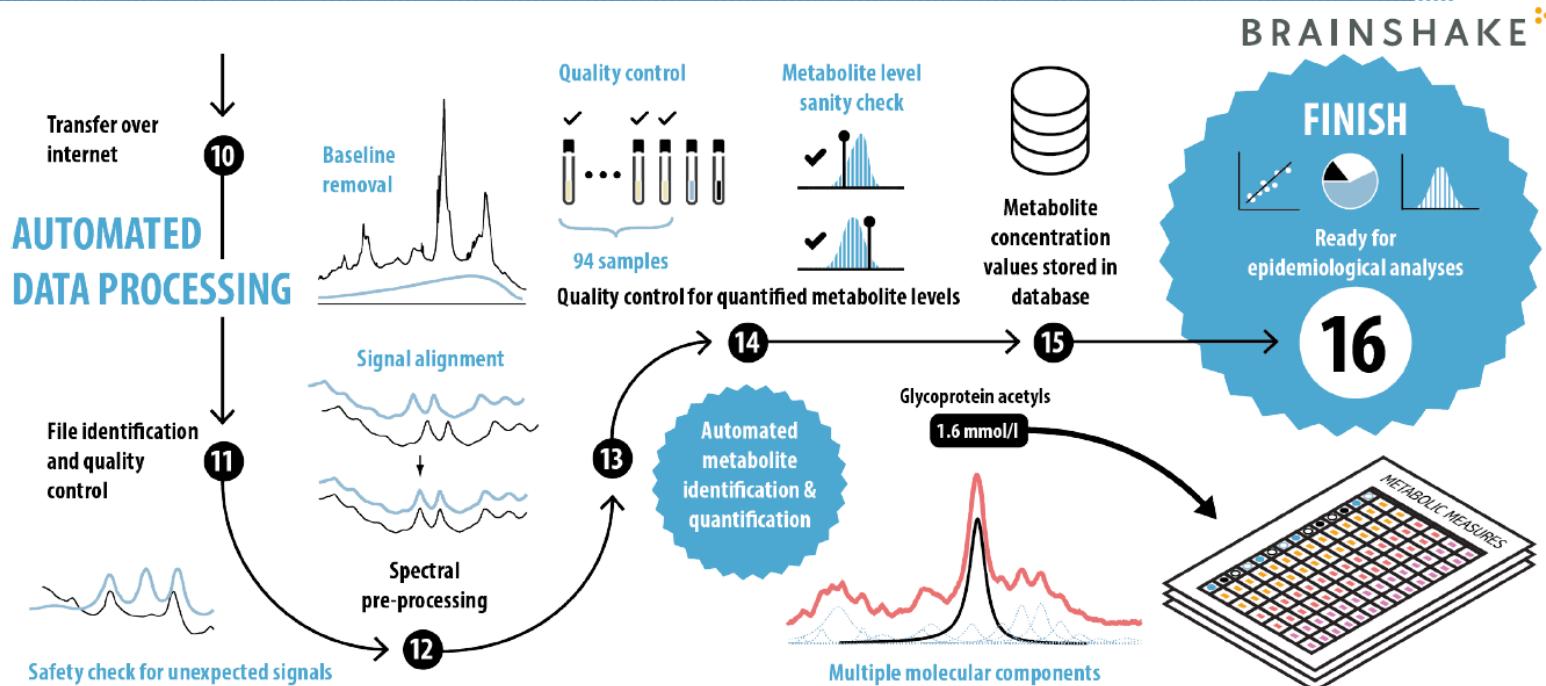
Nightingale



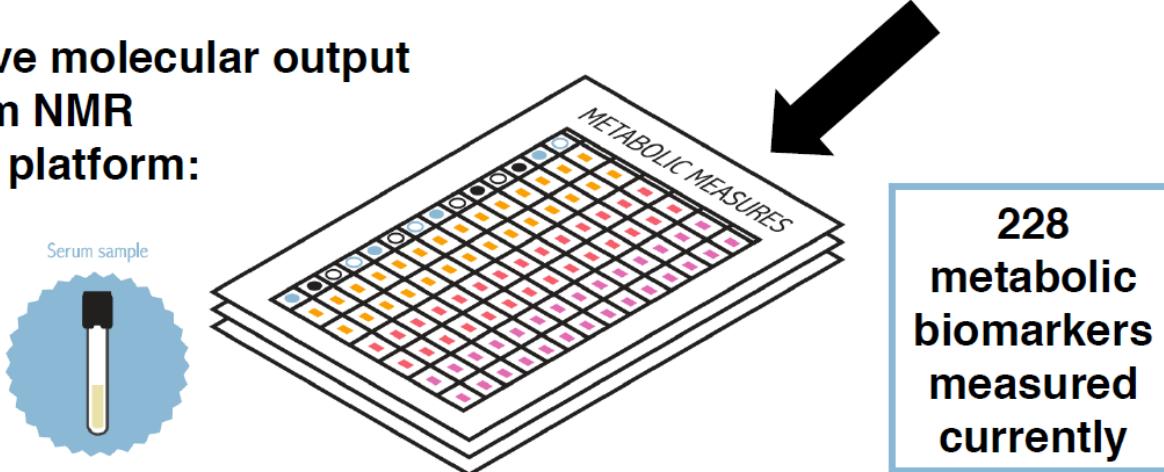
Nightingale Health (formerly known as Brainshake)



Nightingale Health (formerly known as Brainshake)

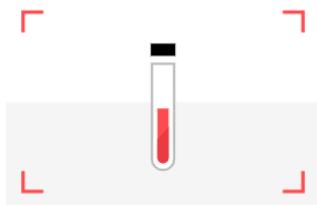


The quantitative molecular output from the serum NMR metabolomics platform:



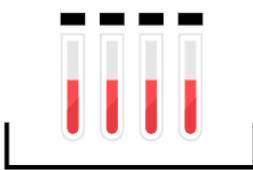
Nightingale Health

Step 1



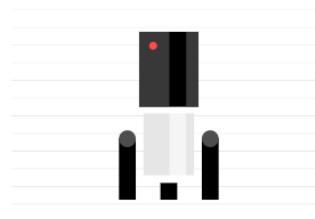
Each sample is tagged for identification.

Step 2



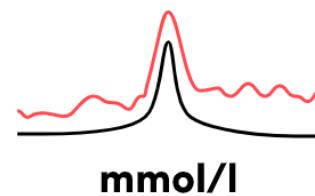
Minimal sample preparation, only buffer added.

Step 3



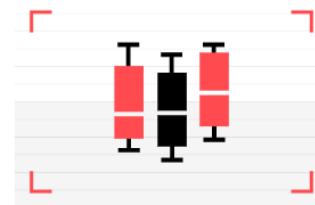
The NMR instrument scans the samples in a non-destructive way.

Step 4



NMR data is transformed into biomarkers of absolute concentrations.

Step 5



Results ready for analysis and interpretation.



Nightingale Health (formerly known as Brainshake)

228 METABOLIC MEASURES

BRAINSHAKE[®]

Ketone bodies (mmol/l)

- Acetate
- Acetoacetate
- 3-hydroxybutyrate

Glycolysis related metabolites (mmol/l)

- Glucose
- Lactate
- Pyruvate
- Citrate
- Glycerol

Inflammation (mmol/l)

- Glycoprotein acetyls

Fatty acids and saturation

- Total fatty acids
- Estimated degree of unsaturation

Fatty acids (mmol/l and % of total FAs)

- Omega-3 fatty acids
- Omega-6 fatty acids
- Polyunsaturated fatty acids
- Monounsaturated fatty acids; 16:1, 18:1
- Saturated fatty acids
- Docosahexaenoic acid; 22:6
- Linoleic acid; 18:2



Apolipoproteins (g/l)

- ApoA-I
- ApoB
- ApoB/ApoA-I

Amino acids (mmol/l)

- Alanine
- Glutamine
- Glycine
- Histidine

Branched-chain amino acids

- Isoleucine
- Leucine
- Valine

Aromatic amino acids

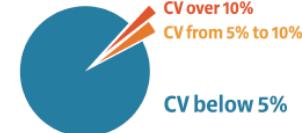
- Phenylalanine
- Tyrosine

Cholesterol (mmol/l)

- VLDL cholesterol
- LDL cholesterol
- HDL cholesterol
- HDL₂ cholesterol
- HDL₃ cholesterol
- Cholesterol
- Free cholesterol
- Esterified cholesterol
- Remnant cholesterol

Fluid balance

- Creatinine (mmol/l)
- Albumin (signal area)



Glycerides & phospholipids (mmol/l)

- VLDL triglycerides
- LDL triglycerides
- HDL triglycerides
- Triglycerides
- Phosphoglycerides
- Ratio of triglycerides to phosphoglycerides
- Phosphatidylcholine and other cholines
- Sphingomyelins
- Total cholines

Lipoprotein particle size (nm)

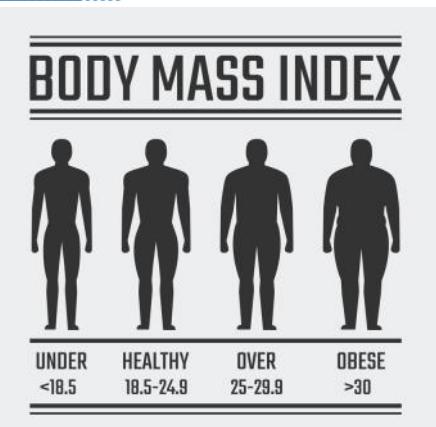
- Mean diameter of VLDL particles
- Mean diameter of LDL particles
- Mean diameter of HDL particles

14 LIPOPROTEIN SUBCLASSES

12 lipid measures for each subclass

- Esterified cholesterol (mmol/l and % of total lipids)
- Free cholesterol (mmol/l and % of total lipids)
- Triglycerides (mmol/l and % of total lipids)
- Phospholipids (mmol/l and % of total lipids)
- Total cholesterol (mmol/l and % of total lipids)
- Total lipids (mmol/l)
- Particle concentration ($\mu\text{mol/l}$)

Metabolomics association with disease



Metabolomics



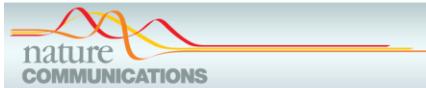
Metabolomics association with disease



Epidemiology and Prevention

Metabolite Profiling and Cardiovascular Event Risk A Prospective Study of 3 Population-Based Cohorts

Peter Würz, PhD; Aki S. Havulinna, DScTech; Pasi Soininen, PhD; Tuulia Tynkkynen, PhD; David Prieto-Merino, PhD; Therese Tillin, MBBS; Anahita Ghorbani, MD; Anna Artati, PhD; Qin Wang, MSc; Mika Tiainen, PhD; Antti J. Kangas, MSc; Johannes Kettunen, PhD; Jari Kaakkonen, MSc; Vera Mikkilä, PhD; Antti Jula, MD, PhD; Mika Kähönen, MD, PhD; Terho Lehtimäki, MD, PhD; Debbie A. Lawlor, MD, PhD; Tom R. Gaunt, PhD



ARTICLE

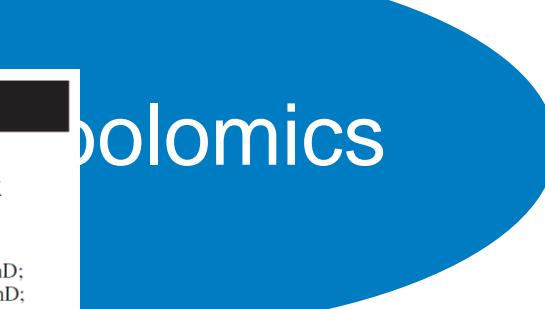
Received 3 Jun 2015 | Accepted 24 Feb 2016 | Published 23 Mar 2016

DOI: 10.1038/ncomms11122

OPEN

Genome-wide study for circulating metabolites identifies 62 loci and reveals novel systemic effects of LPA

Johannes Kettunen et al.[#]



METABOLISM CLINICAL AND EXPERIMENTAL 65 (2016) 111–121



Available online at www.sciencedirect.com

Metabolism

www.metabolismjournal.com



Abdominal obesity and circulating metabolites: A twin study approach

Leonie H. Bogl^{a,*}, Sanna M. Kaye^{b,1}, Joel T. Rämö^{b,c}, Antti J. Kangas^{d,e}, Pasi Soininen^{d,e}, Antti Hakkarainen^f, Jesper Lundbom^f, Nina Lundbom^f, Alfredo Ortega-Alonso^c, Aila Rissanen^{b,g}, Mika Ala-Korpela^{d,e,h,i}, Jaakko Kaprio^{a,c,j,2}, Kirsi H. Pietiläinen^{b,c,k,2}



Metabolomics association with disease



Metabolomics

ORIGINAL ARTICLE

Metabolic Signatures of Insulin Resistance in 7,098 Young Adults

Peter Würtz,^{1,2,3} Ville-Petteri Mäkinen,^{1,4,5} Pasi Soininen,^{1,6} Antti J. Kangas,¹ Taru Tukiainen,^{1,2} Johannes Kettunen,^{2,7} Markku J. Savolainen,^{1,8} Tuija Tammelin,⁹ Jorma S. Viikari,¹⁰ Tapani Rönnemaa,¹⁰ Mika Kähönen,¹¹ Terho Lehtimäki,¹² Samuli Ripatti,^{2,7,13} Olli T. Raitakari,^{14,15} Marjo-Riitta Järvelin,^{3,16,17,18} and Mika Ala-Korpela^{1,6,8}

Pathophysiology/Complications
ORIGINAL ARTICLE

Circulating Metabolite Predictors of Glycemia in Middle-Aged Men and Women

PETER WÜRTZ, PHD^{1,2,3}
MIKA TIANINEN, MSC^{1,4}
VILLE-PETTERI MAKINEN, DSC^{1,5}
ANTTI J. KANGAS, MSC^{1,6}
PASI SOININEN, PHD^{1,7}
JUHA SALTEVO, MD, PHD⁶
SIRKKA KEINANEN-KUKAANNIEMI, MD, PHD^{7,8}

PEKKA MANTYSELKA, MD, PHD^{9,10}
TERHO LEHTIMÄKI, MD, PHD¹¹
MARKKU LAASKO, MD, PHD¹³
ANTTI JULA, MD, PHD¹³
MIKA KÄHÖNEN, MD, PHD¹⁴
MAUNO VANHALA, MD, PHD^{10,15}
MIKA ALA-KORPELA, PHD^{1,3,4,16}

OBJECTIVE—Metabolite predictors of deteriorating glucose tolerance may elucidate the pathogenesis of type 2 diabetes. We investigated associations of circulating metabolites from high-throughput profiling with fasting and postload glycemia cross-sectionally and prospectively on the population level.

Pathophysiology/Complications
ORIGINAL ARTICLE

Branched-Chain and Aromatic Amino Acids Are Predictors of Insulin Resistance in Young Adults

PETER WÜRTZ, PHD^{1,2}
PASI SOININEN, PHD^{2,3}
ANTTI J. KANGAS, MSC²
TAPANI RÖNNEMAA, MD, PHD⁴
TERHO LEHTIMÄKI, MD, PHD⁵

MIKA KÄHÖNEN, MD, PHD⁶
JORMA S. VIIKARI, MD, PHD⁴
OLLI T. RAITAKARI, MD, PHD^{7,8}
MIKA ALA-KORPELA, PHD^{2,3}

Furthermore, the circulating concentrations of branched-chain amino acids (isoleucine, leucine, and valine) and aromatic amino acids (phenylalanine and tyrosine) were recently shown to be associated with the risk of future hyperglycemia and overt



Metabolomics association with disease



Metabolomics



OPEN ACCESS Freely available online

PLOS MEDICINE

Biomarker Profiling by Nuclear Magnetic Resonance Spectroscopy for the Prediction of All-Cause Mortality: An Observational Study of 17,345 Persons

Krista Fischer^{1,*}, Johannes Kettunen^{2,3,4*}, Peter Würtz^{2,4*}, Toomas Haller¹, Aki S. Havulinna³, Antti J. Kangas⁴, Pasi Soininen^{4,5}, Tõnu Esko^{1,6,7,8,9,10}, Mari-Liis Tammesoo¹, Reedik Mägi¹, Steven Smit¹, Aarno Palotie^{2,6,11}, Samuli Ripatti^{2,11}, Veikko Salomaa³, Mika Ala-Korpela^{4,5,12†}, Markus Perola^{1,2‡}, Andres Metspalu^{1,13§}

ARTICLE

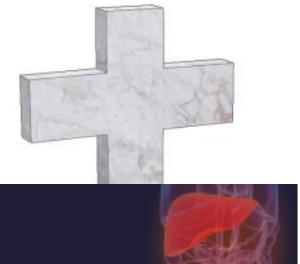
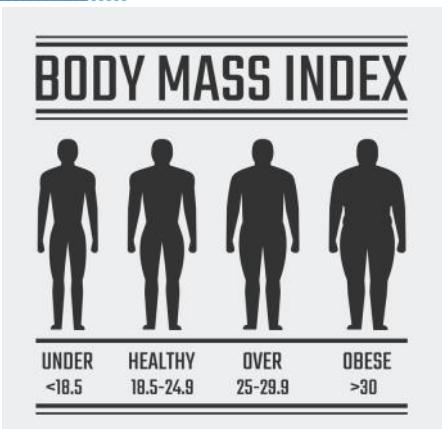
<https://doi.org/10.1371/journal.pmed.1014670>

OPEN

A metabolic profile of all-cause mortality risk identified in an observational study of 44,168 individuals



Metabolomics association with disease



Liver International ISSN 1478-3223

Metabolomics

Liver
INTERNATIONAL

NAFLD/NASH

Ketone body production is differentially altered in steatosis and non-alcoholic steatohepatitis in obese humans

Ville T. Männistö¹, Marko Simonen¹, Jenni Hyysalo², Pasi Soininen^{3,4}, Antti J. Kangas^{3,4}, Dorota Kaminska⁵, Ananda K. Matte⁵, Sari Venesmaa⁶, Pirjo Käkelä⁶, Vesa Kärjä⁷, Johanna Arola⁸, Helena Gylling^{5,9}, Henna Cederberg¹, Johanna Kuusisto¹, Markku Laakso¹, Hannele Yki-Järvinen², Mika Ala-Korpela^{3,4,10} and Jussi Pihlajamäki^{5,11}

The Biomarker GlycA Is Associated with Chronic Inflammation and Predicts Long-Term Risk of Severe Infection

Scott C. Ritchie,^{1,2} Peter Würtz,⁴ Artika P. Nath,^{1,3} Gad Abraham,^{1,2} Aki S. Havulinna,^{5,6} Liam G. Fearnley,^{1,2} Antti-Pekka Sarin,⁶ Antti J. Kangas,⁴ Pasi Soininen,^{4,7} Kristiina Aalto,⁸ Ilkka Seppälä,⁹ Emma Raitoharju,⁹ Marko Salmi,^{5,8} Mikael Maksimow,^{5,8} Satu Männistö,⁵ Mika Kähönen,¹⁰ Markus Juonala,^{11,12} Samuli Ripatti,^{6,13} Terho Lehtimäki,⁹ Sirpa Jalkanen,⁸ Markus Perola,^{5,6} Olli Raitakari,^{14,15} Veikko Salomaa,⁵ Mika Ala-Korpela,^{4,7,16,17,18} Johannes Kettunen,^{4,5,7,19,*} and Michael Inouye^{1,2,3,19,*}

Cell Systems
Report

Learning goals

After todays lectures and practical you

- 1) Can explain the different levels of omics data that are used for disease prediction
 - *Genetics, epigenetics, transcriptomics, proteomics, metabolomics*
- 2) Can explain the strength of the use of metabolomics in disease prediction
 - *Metabolites represent downstream biochemical end products that are closer to the phenotype, because they are also reflect environmental factors*
- 3) Can explain the weaknesses of the use of metabolomics in disease prediction
 - *Snapshot in time and tissue (maybe it was not the best moment or tissue to investigate)*



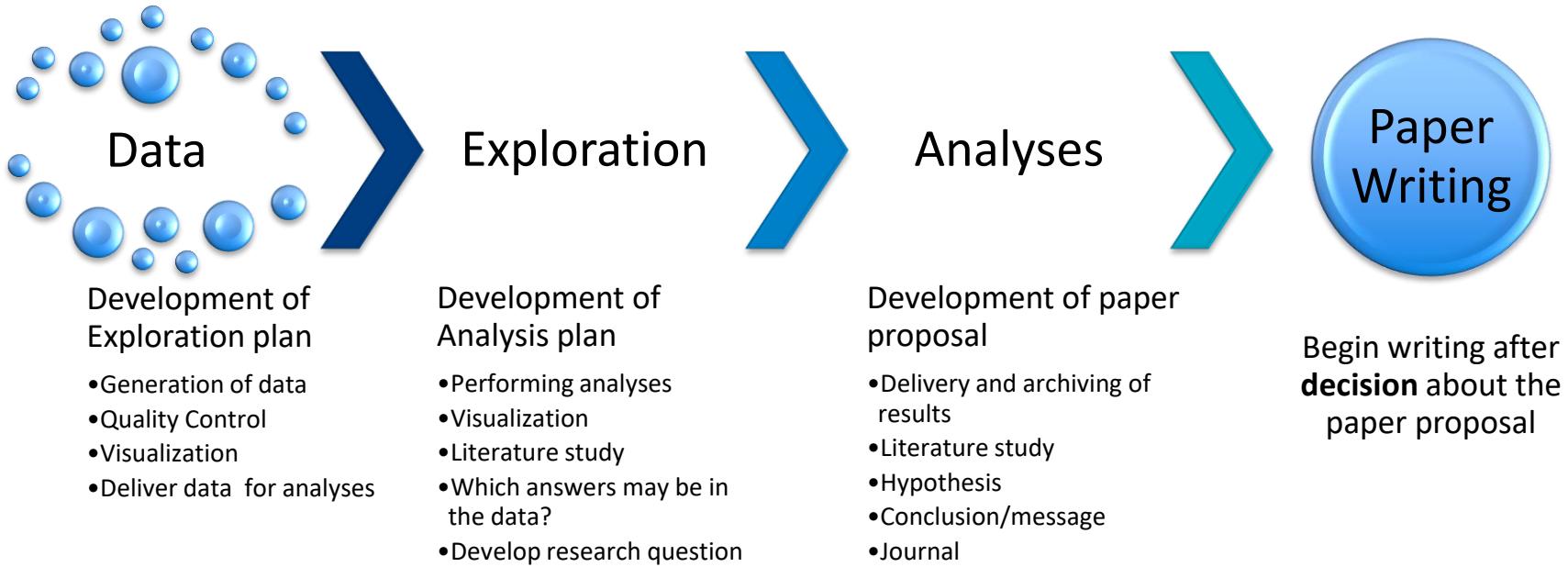
Learning goals

After todays lectures and practical you

- 1) Can explain the different levels of omics data that are used for disease prediction
- 2) Can explain the strength of the use of metabolomics in disease prediction
- 3) Can explain the weaknesses of the use of metabolomics in disease prediction
- 4) Understand the workflow of -omics data analysis
- 5) Explain relevant steps in quality control of metabolomics
- 6) Know which databases may contribute to the interpretation of results



Omics data analysis workflow



Data

1. Sample size
 1. How many subjects do I have?
 2. Number of men, women?
 3. What is the mean age of the subjects?
2. Is the range of measures in the phenotype as expected?



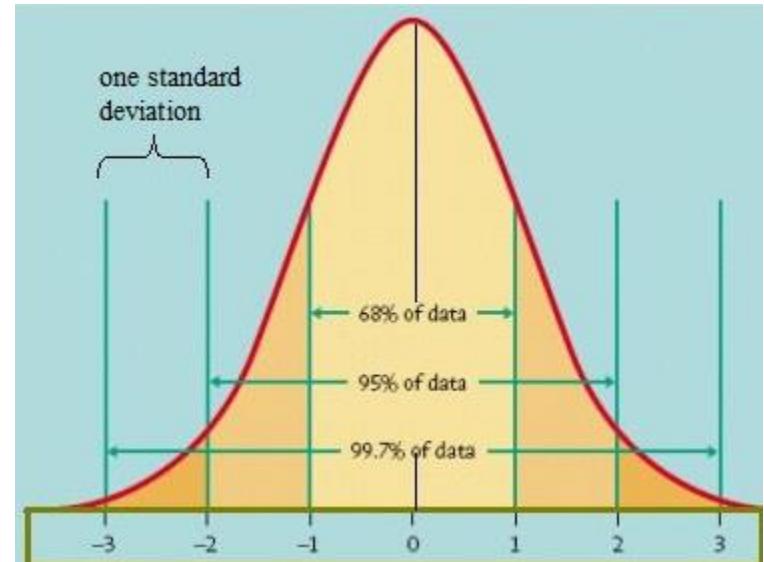
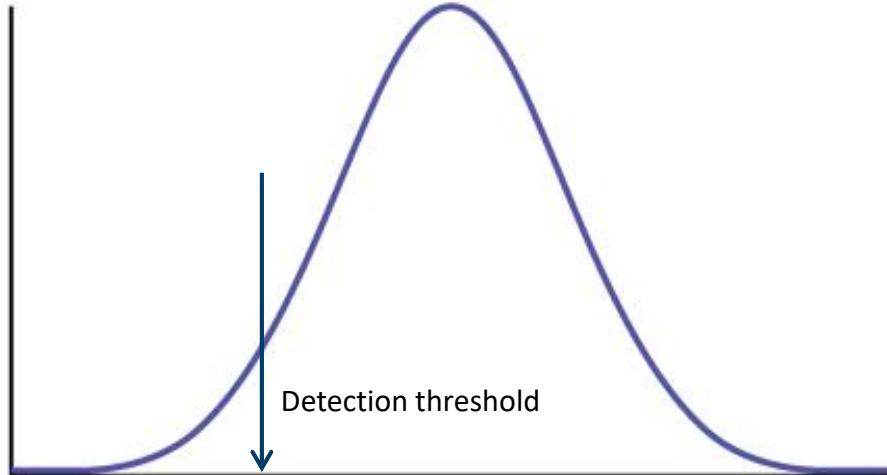
Quality Control: Subject level

1. Are there subject with many missing values?
2. Are there subjects with QC flags from the measurement?
 - Decide **before analysis** what to do with such data!



Quality Control: Metabolite level

1. Are there metabolites missing for many subjects?
2. Are the metabolites with many zero-values?
 - Decide **before analysis** what to do with such data!
3. Are the distributions of all metabolites normal?



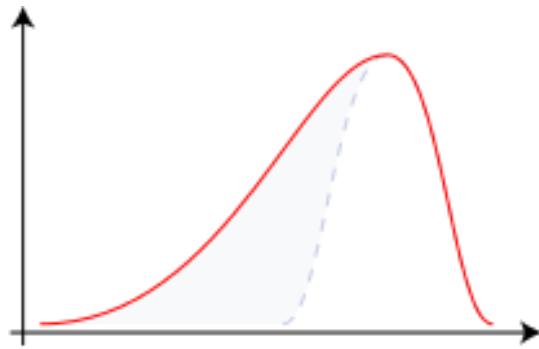
Quality Control: Metabolite level

1. Are there metabolites missing for many subjects?
2. Are the metabolites with many zero-values?
 - Decide **before analysis** what to do with such data!
3. Are the distributions of all metabolites normal?

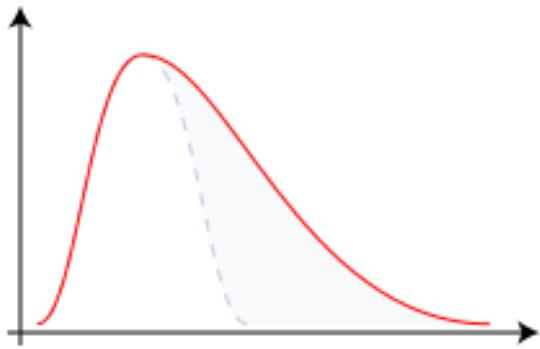
Metabolite level (Y) = Constant + ($\beta_1 \times \text{Sex}$) + ($\beta_2 \times \text{Age}$)



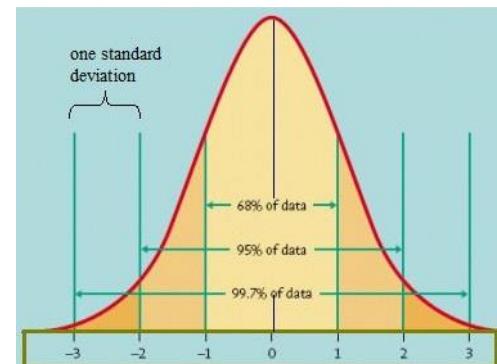
N.B. The outcome of linear regression is assumed to have a normal distribution!



Negative Skew



Positive Skew



Statistical analysis

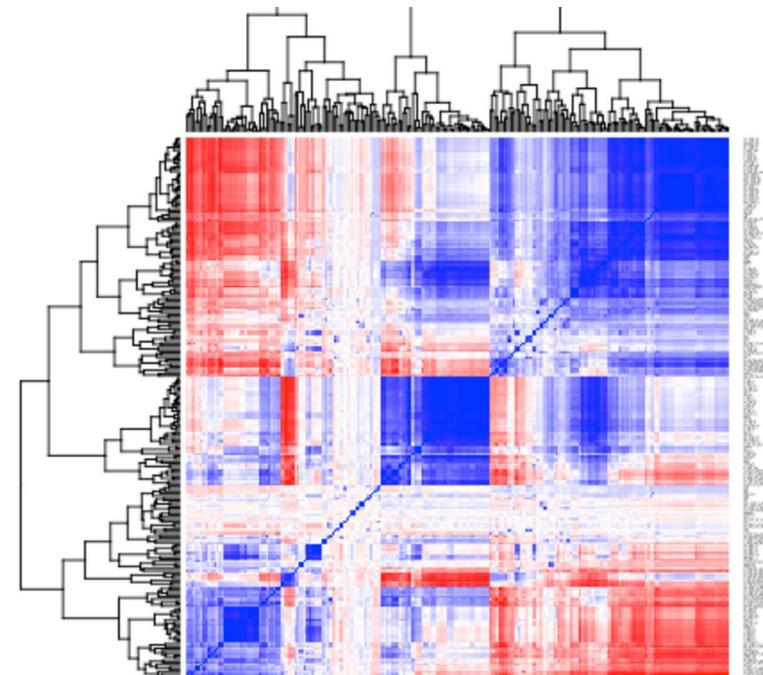
1. Linear regression

$$\text{Metabolite level (Y)} = \text{Constant} + (\beta_1 \times \text{Sex}) + (\beta_2 \times \text{Age})$$

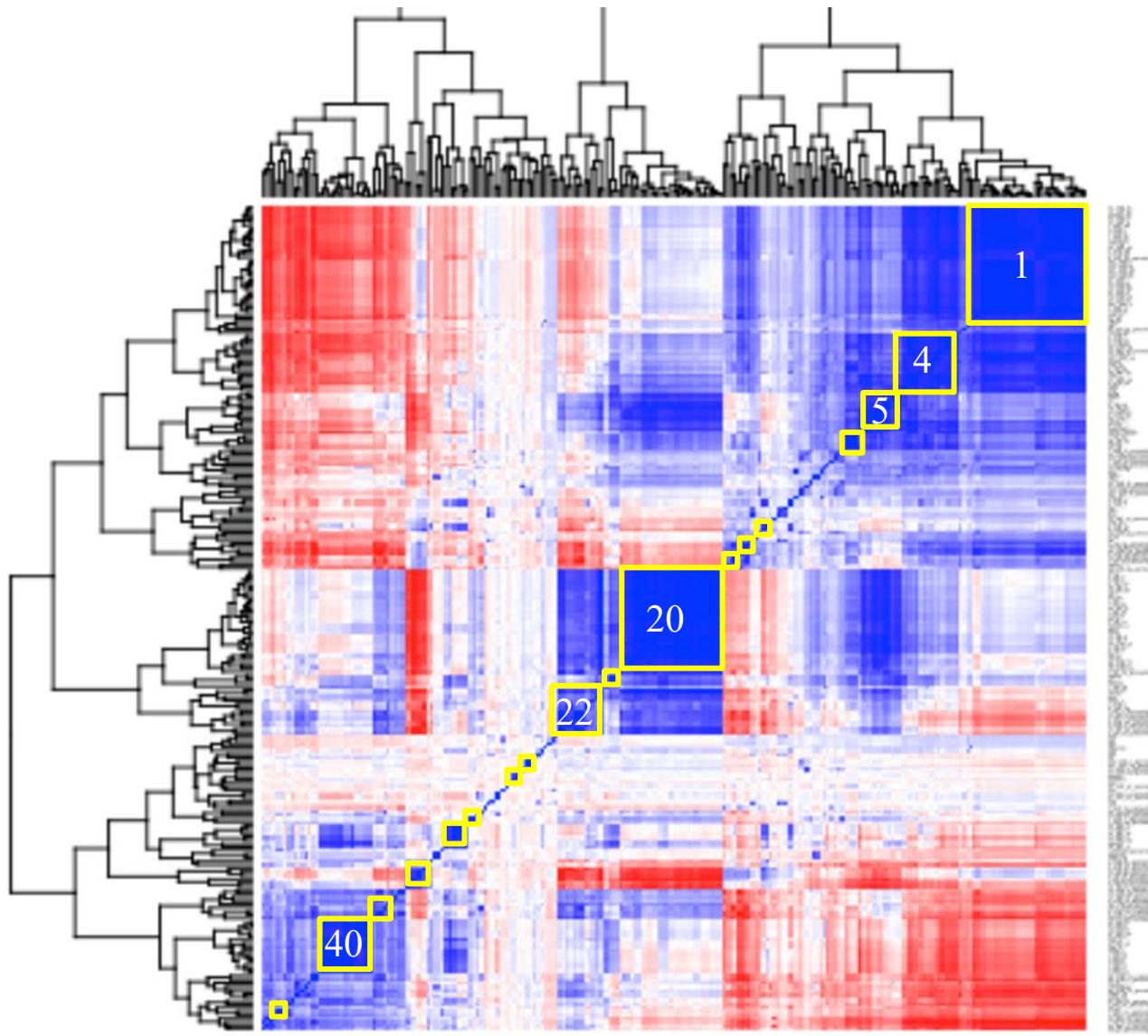
1. How large is effect size (β_1)?
2. What is the significance threshold?

Bonferroni: $p\text{-value} = 0.05/\text{number of independent tests}$

How many independent tests?! ←

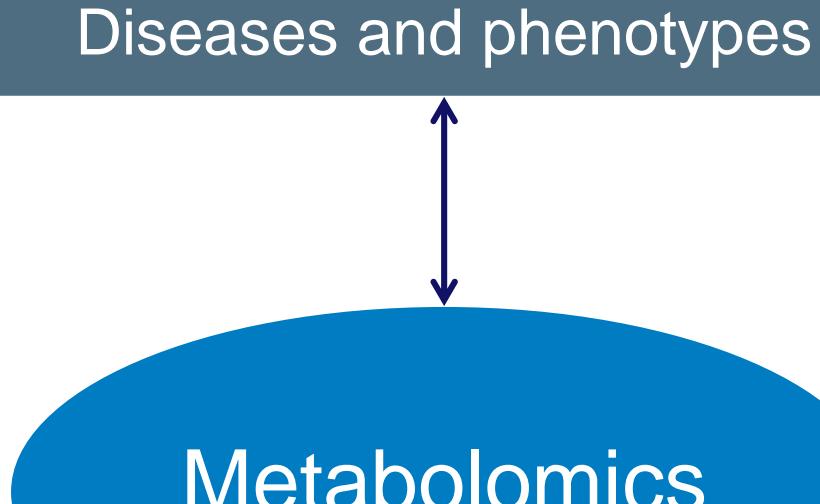


Number of independent tests



~45 clusters:
Significance
threshold =
 $0.05/45 = 0.001$

Metabolomics analyses and interpretation



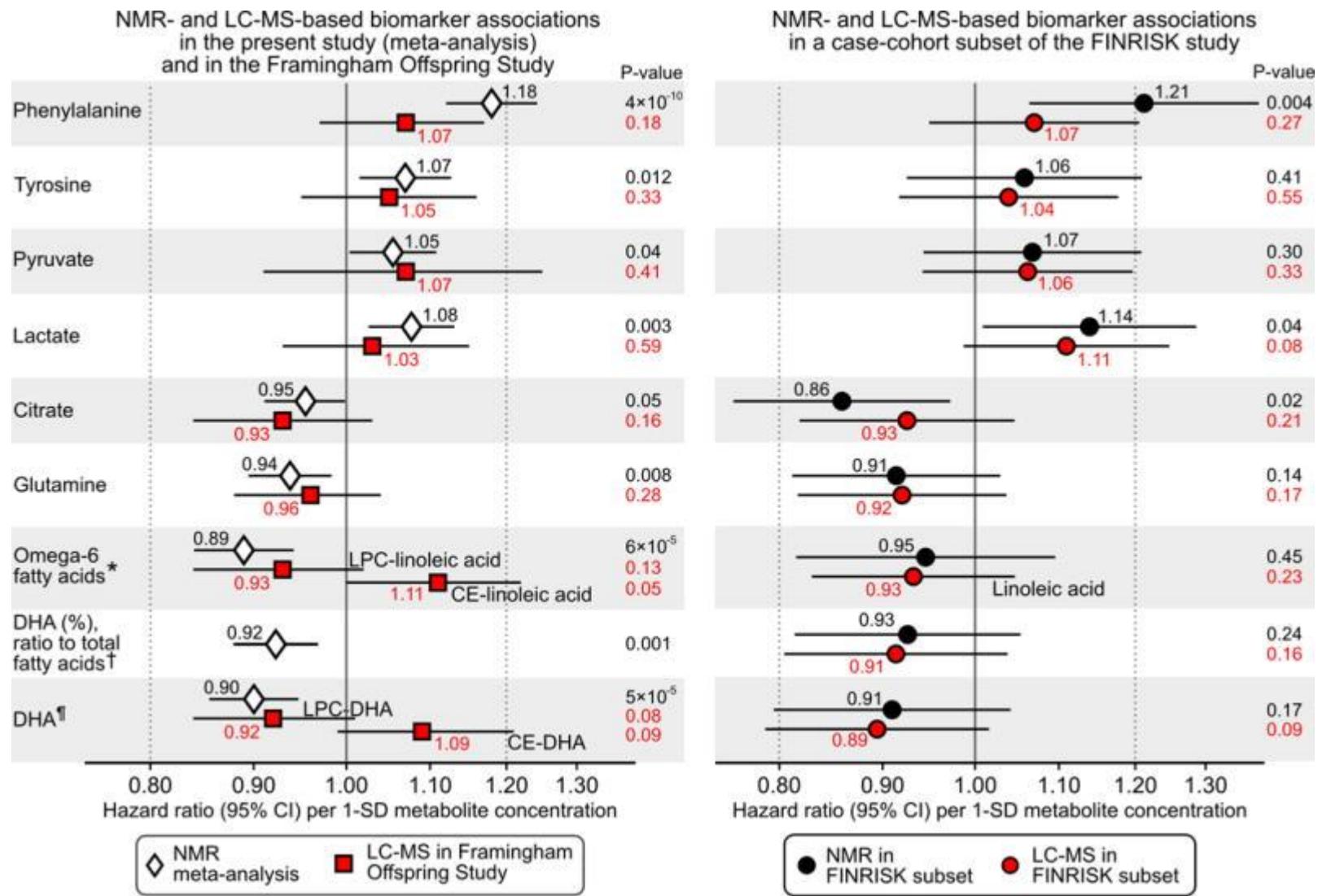
Cross-sectional: Linear regression

$$Y \text{ (Metabolite concentrations)} = C + (\beta_1 \times \text{Disease}) + (\beta_2 \times \text{Age}) + (\beta_3 \times \text{Sex})$$

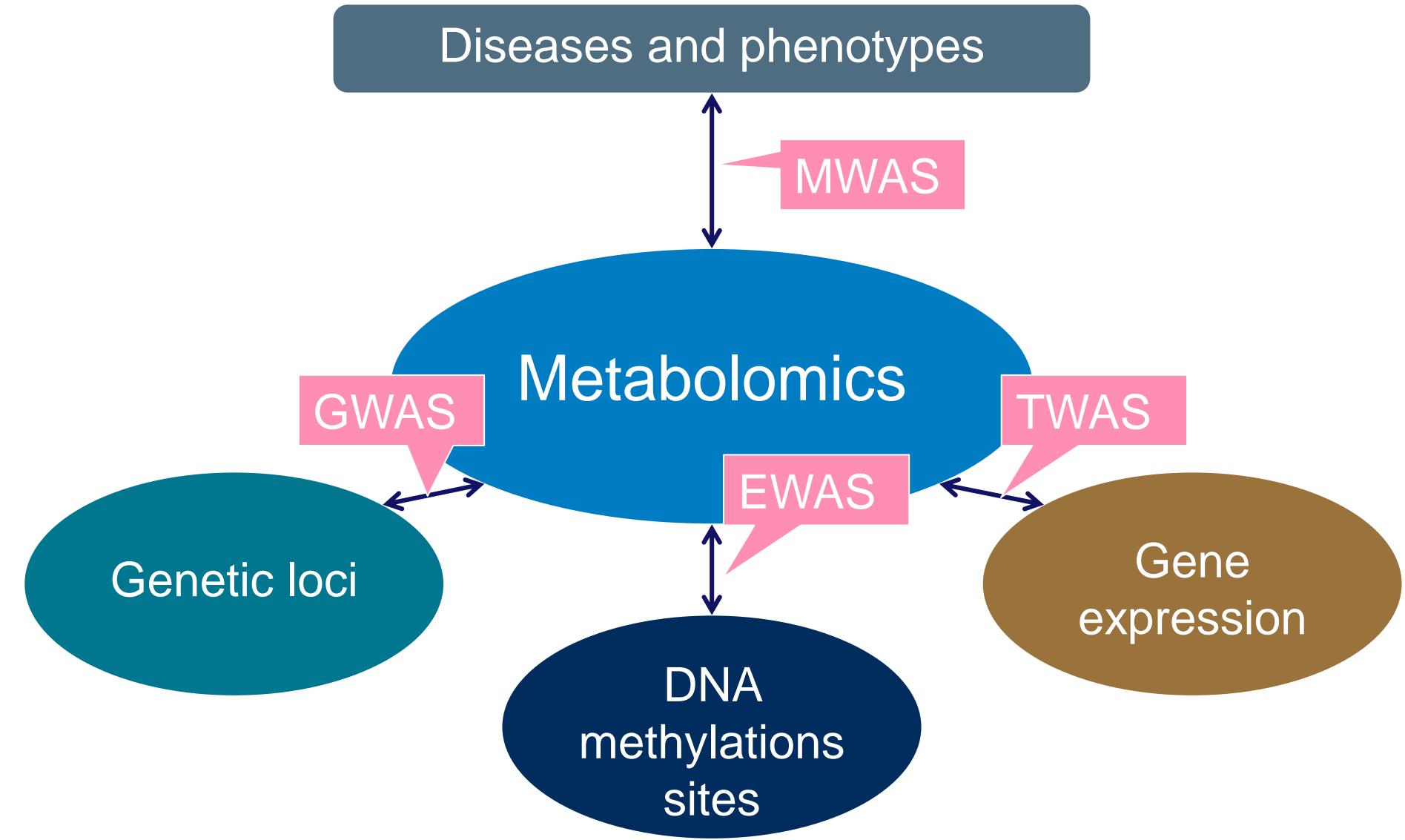
Prospective/Prediction: Cox- regression

$$\ln(H(t)/H_0(t)) \text{ (Hazard for disease)} = (\beta_1 \times \text{Metabolite conc}) + (\beta_2 \times \text{Age}) + (\beta_3 \times \text{Sex})$$

Metabolomics analyses and interpretation



Metabolomics analyses and interpretation



GWAS: genetic loci associating with metabolomics

ARTICLES

nature
genetics

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

Johannes Kettunen^{1,2,37}, Taru Tukiainen^{1,3–5,37}, Antti-Pekka Sarin^{1,2}, Alfredo Ortega-Alonso^{1,6}, Emmi Tikkanen^{1,2}, Leo-Pekka Lyytikäinen⁷, Antti J Kangas⁵, Pasi Soininen^{5,8}, Peter Würtz^{1,3,5}, Kaisa Silander^{1,2}, Danielle M Dick⁹, Richard J Rose^{6,10}, Markku J Savolainen^{11,12}, Jorma Viikari¹³, Mika Kähönen¹⁴, Terho Lehtimäki⁷, Kirsi H Pietiläinen^{1,15,16}, Michael Inouye^{17,18}, Mark I McCarthy^{19,20}, Antti Jula², Johan Eriksson^{21–24}, Olli T Raitakari^{25,26}, Veikko Salomaa², Jaakko Kaprio^{1,6,27}, Marjo-Riitta Järvelin^{3,12,28–30}, Leena Peltonen³⁶, Markus Perola^{1,2,31}, Nelson B Freimer³², Mika Ala-Korpela^{5,8,11,12}, Aarno Palotie^{1,33–35} & Samuli Ripatti^{1,2,33}

med.



ARTICLE

Received 3 Jun 2015 | Accepted 24 Feb 2016 | Published 23 Mar 2016

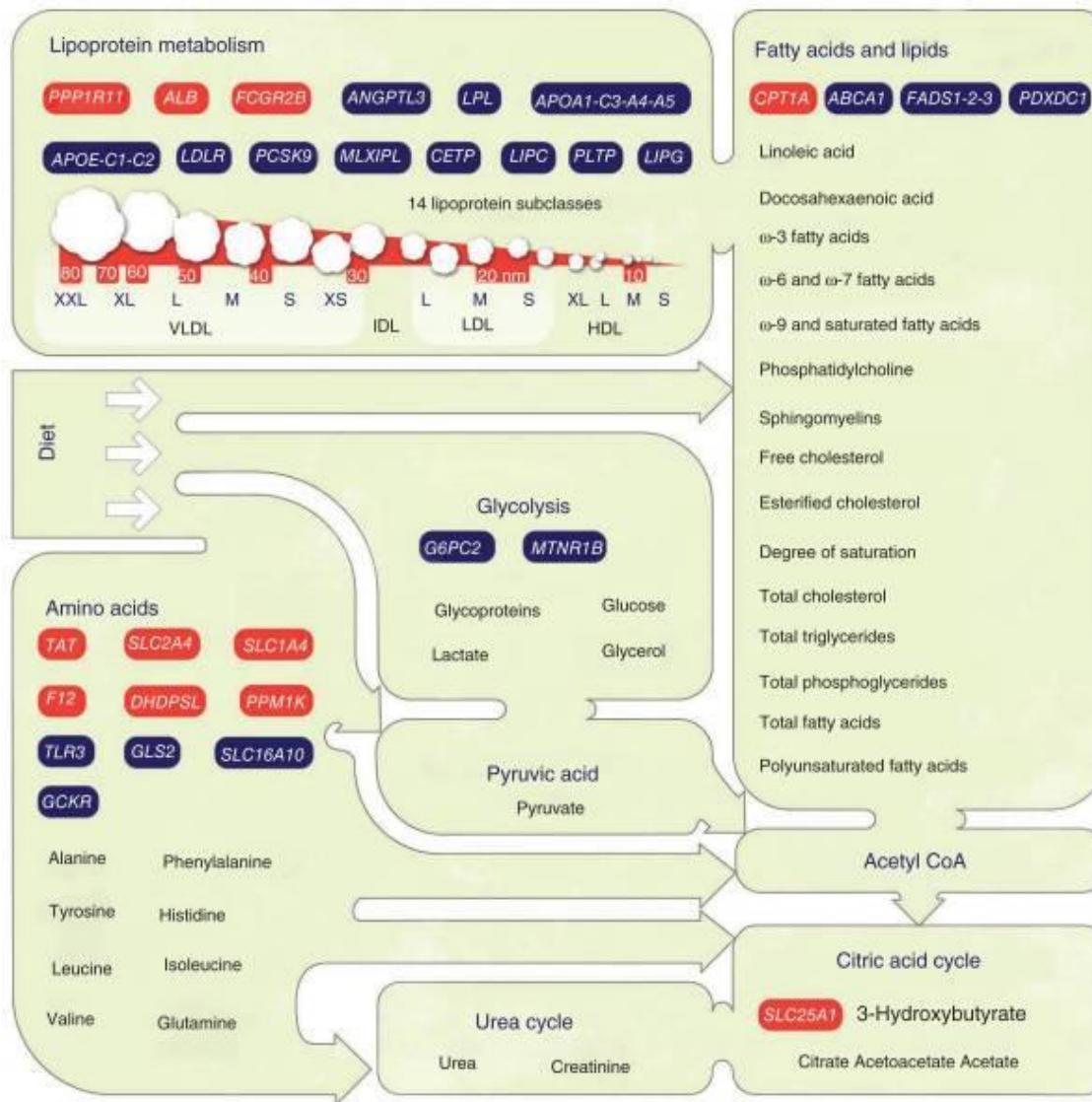
DOI: 10.1038/ncomms11122

OPEN

Genome-wide study for circulating metabolites identifies 62 loci and reveals novel systemic effects of *LPA*

Johannes Kettunen *et al.*#

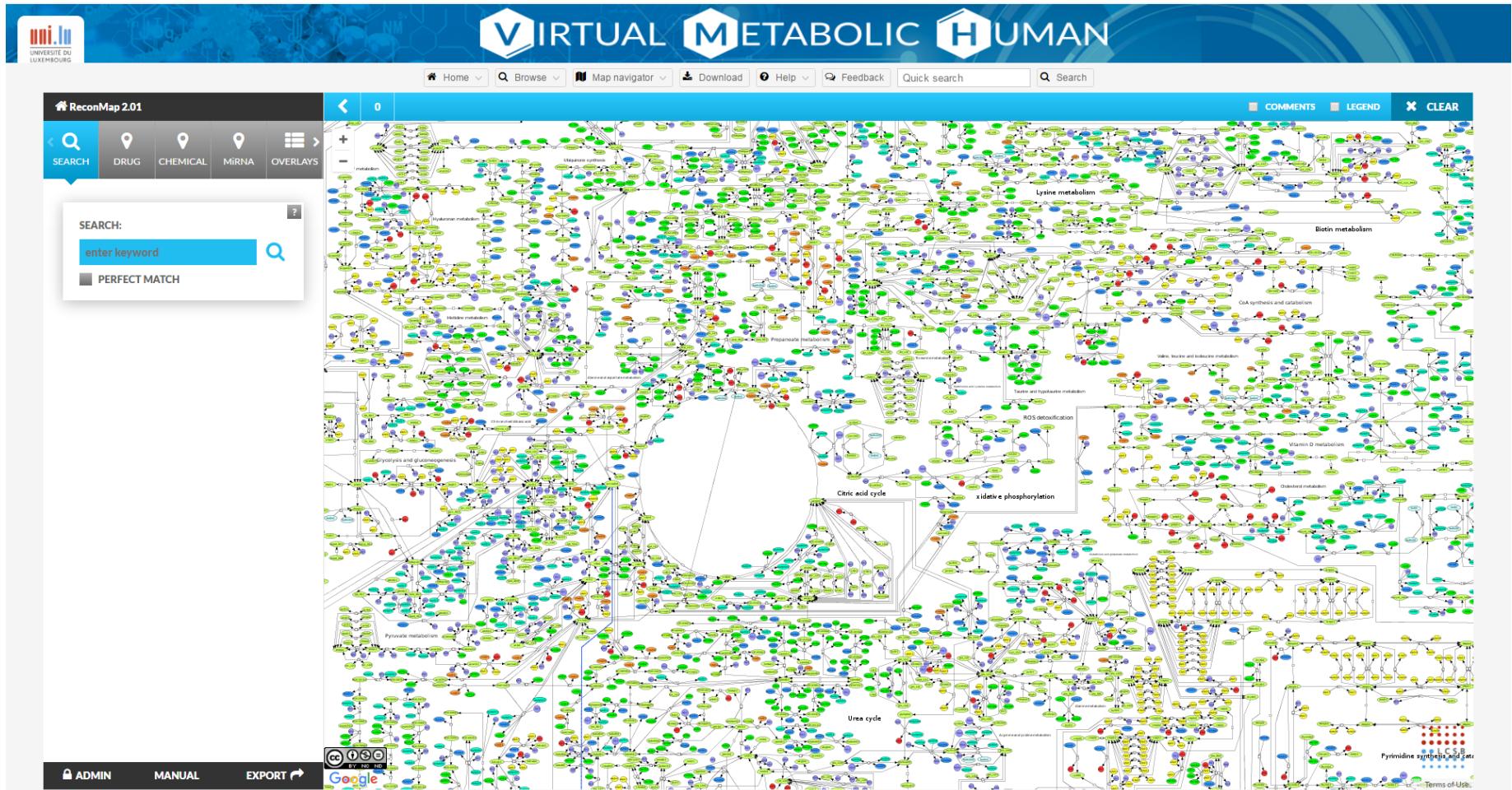
GWAS: genetic loci associating with metabolomics



Virtual Metabolome Human

L U
M C

<http://vmh.uni.lu/#reconmap>



Interpretation and building hypothesis

<http://vmh.uni.lu/#reconmap>

<https://www.ebi.ac.uk/gwas/>

<https://genome.ucsc.edu/>

<https://www.ncbi.nlm.nih.gov/omim>



GWAS Catalog
The NHGRI-EBI Catalog of published genome-wide association studies
 Search the catalog
Examples: breast cancer, rs7329174, Yang, 2q37.1, HBS1L, 6:16000000-25000000



Genome Browser



OMIM

OMIM is a comprehensive, authoritative compendium of human genes and genetic phenotypes that is freely available and updated daily. OMIM is authored and edited at the McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, under the direction of Dr. Ada Hamosh. Its official home is omim.org.

Learning goals

After todays lectures and practical you

- 4) Understand the workflow of -omics data analysis
 - Data exploration, Quality Control, Analysis, Interpretation
- 5) Explain relevant steps in quality control of metabolomics
 - Subject QC, Metabolite QC, Normalisation
- 6) Know which databases may contribute to the interpretation of results
 - A.o. Virtual Metabolome Human, UCSC, OMIM, GWAS catalog



R-Practicals Metabolomics

Molecular Epidemiology

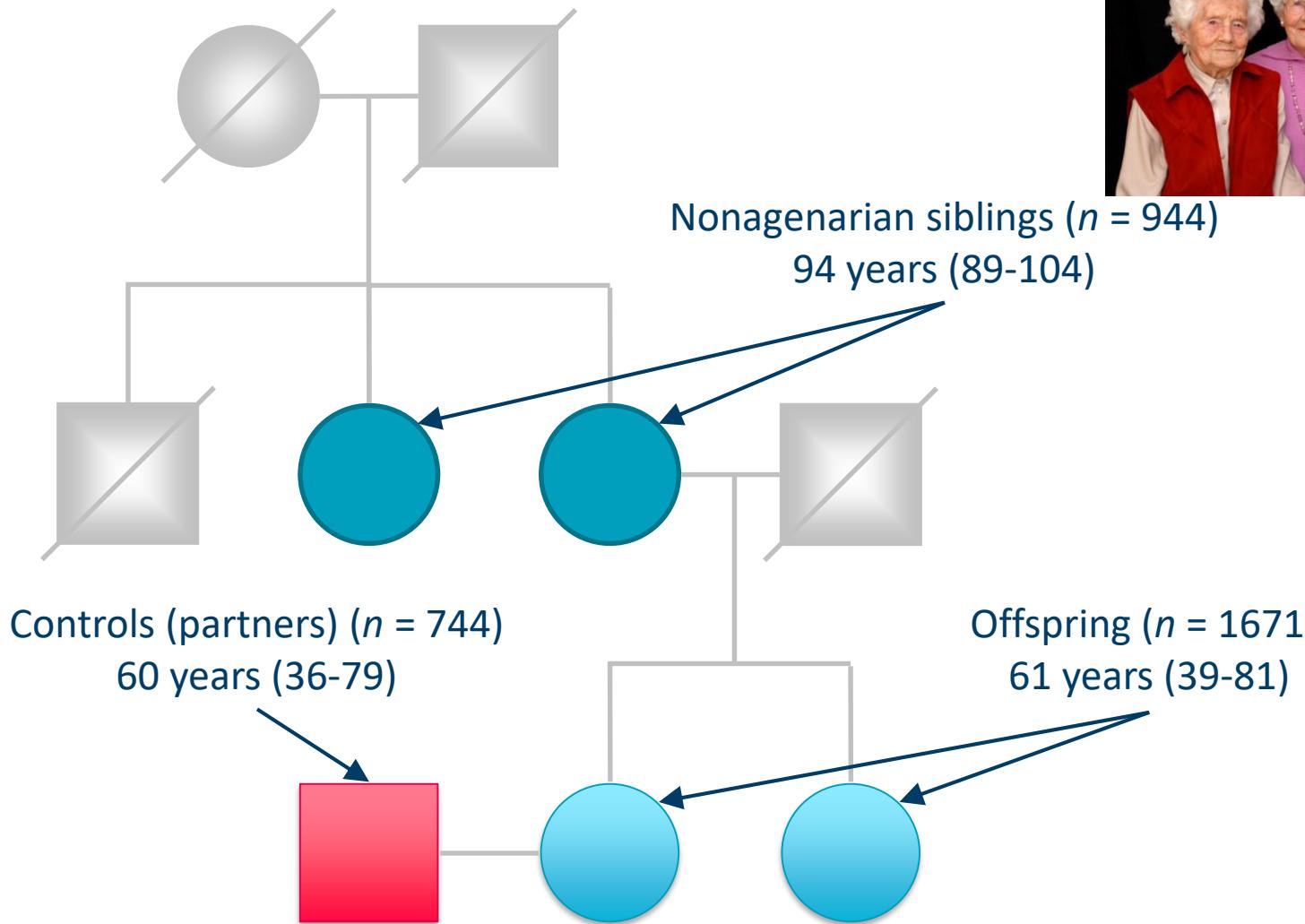
Erik van den Akker

Marian Beekman

Github
Leiden Longevity Study
Non-fasted blood samples
Nightingale platform



Leiden Longevity Study



Practical in R => Erik van den Akker

- 1) Data exploration
- 2) Quality Control
- 3) Data transformation
- 4) Analysis
- 5) Interpretation

© Randy Glasbergen
glasbergen.com



**"Tech support says the problem is located
somewhere between the keyboard and my chair."**