



Skolkovo Institute of Science and Technology

METABOLIC VARIATIONS OF MODERN AND ANCIENT HUMAN
POPULATIONS

Doctoral Thesis

by

VITA STEPANOVA

DOCTORAL PROGRAM IN LIFE SCIENCES

Supervisor
Professor Philipp Khaitovich

Moscow - 2019

© Vita Stepanova 2019

I dedicate this work to my beloved parents Nadezhda Stepanova and Vasily Stepanov who always encouraged me to learn new things and have supported me every single day of my academic journey.

Acknowledgments

I would like to thank my supervisor Philipp Khaitovich for giving me the opportunity to learn from the professionals and to improve my skills. Many thanks to all the lab members: Ekaterina Khrameeva, Ilya Kurochkin, Anna Tkachev, Dmitry Zubkov, Anna Vanyushkina, Pavel Mazin, Waltraud Mair, Elena Stekolschikova and all others for their support, kind attitudes and the many fruitful and productive discussions we had.

I also want to thank our collaborator Svante Pääbo for the inspiration to work on Neanderthal evolution and for sharing his huge experience and knowledge with us.

I also want to express my sincere gratitude to Education Office and all the Skoltech support staff. Without their help and support, I wouldn't have been able to make the progress I did. Their approachability made my work so much easier and I will always be grateful to them.

The professors and academic staff of the Department of Bioengineering and Bioinformatics at Lomonosov Moscow State University deserve a very special mention. Without studying there, I would not be where I am today — thank you for giving me a new mindset and the ability to think critically. In particular, many thanks to Mikhail Gelfand who was my first supervisor and became a significant tutor and mentor both personally and professionally.

I want to thank my family and friends for helping me to not give up and for encouraging me to carry on. You always supported me and were always there for me whenever I needed anything. Sometimes, you even believed in me more than I believed

in myself and that was just what I needed. You pushed me forward and I couldn't have done it without you.

Abstract

Human populations, despite their overwhelming similarity, contain some distinct phenotypic, genetic, epigenetic, and gene expression features. While genetic and gene expression differences among human populations are demonstrated, less is known about differences in the abundance of polar and non-polar low molecular weight compounds, lipids and metabolites. The study provides an evidence of population differences in metabolome and lipidome levels in the cortical region and in the brain and cerebellum. I utilized mass-spectrometry methods to investigate metabolic variations in modern and ancient populations.

For modern populations we assessed the abundance of 1,670 lipids and 258 metabolites in the prefrontal cortex of 146 Han Chinese, 97 Western European, and 60 African American individuals with ages spanning most of lifespan. The statistical analysis and logistic regression models both demonstrate extensive lipid and metabolic divergence of the Han Chinese individuals from the other two populations. This divergence was age-dependent, peaking at approximately 20 years of age, and involved metabolites and lipids clustered in specific metabolic pathways.

For ancient population analysis we first analyzed the metabolomes of humans, chimpanzees and macaques in muscle, kidney and three different regions of the brain. Whereas several compounds in amino acid metabolism occur at either higher or lower concentrations in humans than in the other primates, metabolites in oxidative phosphorylation and purine biosynthesis are consistently present in lower concentrations in the brains of humans. In particular, metabolites downstream of adenylosuccinate lyase,

which catalyzes two reactions in purine synthesis, occur at lower concentrations in humans. This enzyme carries an amino acid substitution unique to modern humans relative to Neanderthals. Secondly, by introducing the modern human substitution into the genomes of mice, as well as the ancestral, Neanderthal-like substitution into the genomes of human cells, we showed that this single amino acid substitution is responsible for much or all of the metabolic changes affecting purine biosynthesis in humans. Thus, at least one substitution that became fixed among humans since their divergence from Neanderthals has consequences for intermediary metabolism.

The results of the current work provide evidence of modern human-specific changes in brain development and highlight the inter-population variations of the molecular phenotype of the brain.

Keywords: *populations, metabolome, lipidome, Neanderthal, ADSL, purine biosynthesis, mass spectrometry, brain.*

Table of Contents

ACKNOWLEDGMENTS	3
ABSTRACT	5
TABLE OF CONTENTS	7
CHAPTER 1. INTRODUCTION	9
ADVANTAGES OF METABOLOMICS AND LIPIDOMICS	9
SIGNIFICANCE OF THE WORK	9
IMPLICATIONS OF THIS WORK	11
PERSONAL CONTRIBUTION	11
PUBLICATIONS	12
CONFERENCES	13
LIST OF SYMBOLS, ABBREVIATIONS	14
CHAPTER 2. REVIEW OF THE LITERATURE	16
HUMAN METABOLOMICS AND LIPIDOMICS	16
<i>Metabolites and lipids of the human brain</i>	18
<i>Metabolite and lipid measurements</i>	19
MODERN HUMAN POPULATION DIVERSITY	20
INTROGRESSION OF ANCIENT HUMAN POPULATION	23
CHAPTER 3. RESULTS	28
DIFFERENCES IN METABOLOME ORGANIZATION IN MODERN POPULATIONS	28
<i>Lipidome and metabolome variation analysis</i>	28
<i>Statistical analysis of lipid and metabolite differences among populations</i>	31
<i>Population classification using machine learning</i>	34
<i>Age dynamics of population difference</i>	35
<i>Functional characterization of HC-specific lipid and metabolite differences</i>	38
DIFFERENCES IN METABOLOME ORGANIZATION IN ANCIENT POPULATIONS	39
<i>Metabolic changes unique to humans</i>	39
<i>Metabolic pathways unique to humans</i>	41
<i>The metabolome of the humanized Adsl mice</i>	42
<i>Purine biosynthesis in the humanized mice</i>	46
<i>Activity and stability of modern humanized mouse ADSL</i>	46
<i>Validation of metabolic changes using cell lines</i>	47
<i>Comparison to chimpanzee cells</i>	48
<i>Purine biosynthesis in autistic metabolome</i>	49
<i>Evolution of purine biosynthesis metabolites and ASD</i>	51
CHAPTER 4. DISCUSSION	54
CHAPTER 5. METHODS	63
MASS SPECTROMETRY METHODS	63
<i>GC-MS measurements</i>	63
<i>CE-MS measurements</i>	63
<i>LC-MS measurements</i>	64
COMPUTATIONAL METHODS	66

<i>Lipid compounds preprocessing</i>	66
<i>Polar compounds preprocessing</i>	67
<i>Data filtration</i>	68
<i>Sample sets definitions</i>	69
<i>T-distributed Stochastic Neighbor Embedding (t-SNE) analysis</i>	69
<i>Population specificity analysis</i>	70
<i>Population analysis within specified age groups</i>	72
<i>Consistency analysis</i>	73
<i>Lipid annotation and enrichment analysis</i>	75
<i>Principal component analysis</i>	75
<i>Differential concentration test</i>	75
<i>Total variation analysis</i>	76
<i>ANCOVA analysis</i>	76
CHAPTER 6. CONCLUSIONS	77
BIBLIOGRAPHY	79

Chapter 1. Introduction

Advantages of metabolomics and lipidomics

Over the last two decades, there has been an explosion in interest in “omics” research and corresponding development in its research techniques and methods. Genomics and transcriptomics were believed to answer a lot of questions regarding the aetiology of disease, but it became clear that, despite their usefulness, they did not provide the whole picture. It has become clear that we need to recruit new technologies to study those factors that cannot be explained by genomics and transcriptomics, such as environment or diet. Metabolomics and lipidomics, which remain largely unexplored in the brain, bring us a step closer to phenotype, and thus could support the observed connection between genomic and phenotypic signatures or even open up new avenues of research. For example, we have sequenced the Neanderthal genome and have studied the genomic differences of coding regions between modern humans and Neanderthals, but the Neanderthal extinction still remains unexplained.

Significance of the work

Lipidome and metabolome organization of the human brain remains largely unexplored. In this work I present a systematic analysis of the lipidome and metabolome organization of the prefrontal cortex gray matter of modern individuals of Han Chinese, Western European, and African American descent. We demonstrate that the lipid and metabolic brain composition differs substantially between the Han Chinese and the other two populations. Furthermore, this difference is age-dependent, peaking in young adults,

and involves metabolites and lipids clustering in specific metabolic pathways. The results represent the first large-scale study reporting the presence of substantial lipid and metabolic brain composition differences among contemporary human populations. While the study design does not allow us to decouple genetic and environmental effects, the reported differences exist in the populations and may be essential in facilitating further studies of differential cognitive disease susceptibility and the optimal treatment strategies in individuals of different descent.

In addition to investigating the metabolic traits of modern individuals, I also study the potential importance of the metabolic traits in the evolution of modern humans. To find metabolic differences that set modern humans apart from their closest evolutionary relatives I investigate the metabolomes of the brain, muscle and kidney in humans, apes and monkeys. I find that steady-state concentrations of many compounds involved in amino acid metabolism are present in higher or lower levels in humans than in other primates. In the future, it may be of interest to investigate the consequences of these human-specific metabolic features for the synthesis and catabolism of amino acids. I also show that purine biosynthesis is decreased in all tissues analyzed, although most drastically in the brain. We hypothesized that the a single amino acid substitution in the purine biosynthesis enzyme has contributed to the reduction seen in modern human tissues and the pronounced down-regulation in the brain. Given that the mutations in humans that affect enzymes involved in purine metabolism have grater pathological consequences in the nervous system than in other organs, the substitution might induce human-specific changes in brain development and function. Future work will have to

address this and other possibilities. Moreover, our results provide the first evidence of the effect of modern human-specific amino acid substitutions.

Implications of this work

The study shows evidence of brain metabolome and lipidome differences between human populations. For modern humans, it could help to consider population specificity when searching for appropriate medical treatment or diet. In evolutionary terms, our observations support the concept of a biochemical level of human brain development.

It has been previously found that there are more than 90 amino acid substitutions between modern humans and Neanderthals. Their combinations remain unstudied but have great potential for the future study of human evolution. We have opened a new field in the study of Neanderthal evolution: construct transgenic mice models or edited cell lines and study the metabolic changes. Our results provide the first evidence of the effect of modern-human-specific amino acid substitutions, demonstrating the effectiveness of this approach for the future investigation of amino acid substitutions.

Personal contribution

The author conducted most of the bioinformatics analysis presented in the thesis. In the study of modern human population differences the author performed statistical analysis and annotation of molecular compounds. The author designed and executed whole bioinformatics analysis of the evolutionary metabolic alterations.

The results and methods of the thesis mention the kinetic, biochemical and genetic experiments were carried out by colleagues.

Publications

1. Tkachev A*, **Stepanova V***, Lei Z*, Khrameeva E, Zubkov D, Giavalisco P, Khaitovich P. Differences in lipidome and metabolome organization of prefrontal cortex among human populations. *Sci Rep [in production]*.

* shared first authors

2. Kurochkin I*, Khrameeva E*, Tkachev A, **Stepanova V**, Vanyushkina A, Li Q, Zubkov D, Shichkova P, Halene T, Willmitzer L, Giavalisco P, Akbarian S, Khaitovich P. Metabolome signature of autism in the human prefrontal cortex. *Commun Biol* 2019 June: 2.

3. Xu C, Li Q, Efimova O, He L, Tatsumoto S, **Stepanova V**, Oishi T, Udono T, Yamaguchi K, Shigenobu S, Kakita A, Nawa H, Khaitovich P, Go Y. Human-specific features of spatial gene expression and regulation in eight brain regions. *Genome Res.* 2018 Aug 28(8):1097-1110.

4. Mazin PV, Shagimardanova E, Kozlova O, Cherkasov A, Sutormin R, **Stepanova VV**, Stupnikov A, Logacheva M, Penin A, Sogame Y, Cornette R, Tokumoto S, Miyata Y, Kikawada T, Gelfand MS, Gusev O. Cooption of heat shock regulatory system for anhydrobiosis in the sleeping chironomid *Polypedilum vanderplanki*. *Proc Natl Acad Sci U S A.* 2018 Mar 6;115(10)

5. Flegontov P, Changmai P, Zidkova A, Logacheva MD, Altınışık NE, Flegontova O, Gelfand MS, Gerasimov ES, Khrameeva EE, Konovalova OP, Neretina T, Nikolsky YV, Starostin G, **Stepanova VV**, Travinsky IV, Tříška M, Tříška P, Tatarinova TV

Genomic study of the Ket: a Paleo-Eskimo-related ethnic group with significant ancient North Eurasian ancestry. *Sci Rep.* 2016 Feb 11;6:20768.

6. Leyn SA, Suvorova IA, Kazakov AE, Ravcheev DA, **Stepanova VV**, Novichkov PS, Rodionov DA. Comparative genomics and evolution of transcriptional regulons in *Proteobacteria*. *Microb Genom.* 2016 Jul 11;2(7).

Conferences

1. Stepanova V., Khrameeva E., Moczulska K., Khaitovich P., Pääbo S. Metabolic changes induced by human-specific amino acid substitution in ADSL protein. The Annual Meeting of the Society of Molecular Biology and Evolution (SMBE), 2017.

2. Stepanova V., Mair W., Pääbo S., Rogaev E., Khaitovich P. Brain metabolic features associated with behavioral traits and domestication. The II Joint Congress on Evolutionary Biology 2018.

List of Symbols, Abbreviations

AA — African American

ADSL — Adenylosuccinate lyase

AIR — 5-Aminoimidazole ribonucleotide

AMP — Adenosine monophosphate

ANCOVA — Analysis of Covariance

ANOVA — Analysis of Variance

ASD — Autism Spectrum Disorder

AUC — Area under the ROC Curve

BH — Benjamini-Hochberg

CB — Cerebellum

CD — Circular Dichroism

CE-MS — Capillary electrophoresis–mass spectrometry

CNS — Central Nervous System

GC-MS — Gas chromatography–mass spectrometry

GMP — Guanosine monophosphate

GSEA — Gene Set Enrichment Analysis

HC — Han Chinese

IMP — Inosine Monophosphate

ITP — Inosine Triphosphate

KEGG — Kyoto Encyclopedia of Genes and Genomes

LC-MS — Liquid chromatography–mass spectrometry

MDS — Multidimensional Scaling

PFC — Prefrontal Cortex

PMI — Post-mortem interval

PVC — Primary Visual Cortex

RIN — RNA integrity number

S-Ado — Succinyladenosine

SAICAr — Succinylaminoimidazole carboxamide riboside

SVR — Support Vector Regression

t-SNE — t-Distributed Stochastic Neighbor Embedding

WE — Western European

wt — Wild type

Chapter 2. Review of the Literature

Human metabolomics and lipidomics

Metabolites, commonly defined as polar compounds with molecular weight below 1500Da, form the molecular phenotype of an organism and include substrates, intermediates or products of biochemical reactions. Together, these metabolites comprise the metabolome of an organism. Human metabolites, predominantly present in blood, have been diagnostic markers for decades (Grant *et al.*, 1970; Psychogios *et al.*, 2011). Routine clinical diagnostic tests measure the concentrations of more than twenty metabolites, including glucose, creatine, urea, and bilirubin. In order to fully catalogue all metabolites in the human body, the Human Metabolome Project (HMP) was initiated in 2004. As a result thousands of metabolites have been identified and quantified in various human tissues and the data has been published in the Human Metabolome Database (HMDB) (Wishart *et al.*, 2018). Currently, the database contains information about 114 100 metabolites and continues to expand actively. Each entry contains the metabolite chemical, clinical, functional, and biochemical data of the metabolite, as well as information about the metabolite performance in different tissues, cell compartments and environmental factors.

Metabolites represent different chemical classes of organic molecules and play a vast variety of functional roles, representing building blocks for larger biological molecules, as well as co-factors, energy molecules, signaling factors, and so on (James, 2016). Furthermore, metabolite functions, as well as concentration levels, vary among the

tissues. For example, adipose tissue releases lactate, whereas liver and kidney utilize it for glucose production (Adeva-Andany *et al.*, 2014).

Lipids, a non-polar fraction of metabolites, also comprise a diverse class of molecules with characteristic chemical and structural properties that distinguish this group from the others. Most commonly, lipids are defined as compounds insoluble in water, but soluble in non-polar organic solvents (“Oxford Dictionary of Biochemistry and Molecular Biology,” 2007). Chemically, the term “lipids” could be classified as fatty-acids derivatives and sterol-containing molecules (Fahy *et al.*, 2005). Lipids are integrated into almost all processes of the human body: they provide substrates for the energy exchange and energy storage, form cellular and organellar membranes, provide the microenvironment for membrane proteins, perform the transport function for hydrophobic and amphiphilic compounds (Green *et al.*, 1966; Vance *et al.*, 2008; Zabrecky *et al.*, 1985). Additionally, some lipids act as hormones and secondary messengers (Dennis *et al.*, 1991).

Lipidome is the term given to describe the entirety of the lipids that exist in a biological sample. In this work, we examine the lipidome of the brain. The lipidome is a constituent of the metabolome, but in this work we consider the two separately. This is because the techniques used to detect non-polar lipids and other metabolites are different, and the distinction between metabolome and lipidome aids clarity.

Alterations in lipid abundance have been associated with a large number of disorders, including cancer (Görke *et al.*, 2010; Min *et al.*, 2011), diabetes (Han *et al.*, 2018), atherosclerosis (Ekroos *et al.*, 2010), hypertension (Graessler *et al.*, 2009) and obesity

(Pietiläinen *et al.*, 2007). The effect of these diseases on lipid abundance means that techniques to detect lipids have the potential to become a diagnostic tool in the future.

Metabolites and lipids of the human brain

The components of amino acid and lipid metabolic pathways account for the largest proportion of the brain metabolome (Bogner *et al.*, 2012; Xiaojiao Zheng *et al.*, 2016). Amino acids play diverse functional roles in the brain. They comprise proteins and perform signaling functions. Nonessential amino acids, such as glutamine and aspartate, act as neurotransmitters in the brain. They mediate excitatory synaptic transmission in neurons (Abarca *et al.*, 1999). Glutamatergic transmission is implicated in neuroplasticity, including memory formation (Abarca *et al.*, 1999).

The brain accounts for approximately 20% of the body's total energy use, and, uses glucose as a sole energy source (Mergenthaler *et al.*, 2013). Glucose is utilized as a fuel for cellular energy through the generation of ATP (Erbslöh *et al.*, 1958). Purines also make a significant contribution to brain metabolome composition (Xiaojiao Zheng *et al.*, 2016). In addition to being a crucial building component of nucleic acid molecules and the energy source, they are also involved in extracellular communication (Lombard, 2006).

Lipids are the most abundant organic molecules of the brain, crucial for membrane architecture and geometry, cell signaling, and protein anchoring (Brady *et al.*, 2012). Lipidome organization of the human brain was shown to be significant for neural development and function (Davletov *et al.*, 2010; Hunter *et al.*, 2018), and to be altered in CNS disorder progression (Abbott *et al.*, 2015; Lukiw, 2005), and in response to medical

treatment (Yu *et al.*, 2014). In addition to non-polar lipids, differences in abundance of metabolites in the prefrontal cortex were linked to neuropsychiatric disorders, including schizophrenia and psychosis, as well as the extent of cognitive abilities, such as memory and orientation (Chen *et al.*, 2014; Crabtree *et al.*, 2018, 2016; Dwyer *et al.*, 2001; Fanfan Zheng *et al.*, 2018). Moreover, the lipid and metabolite composition of the human brain was demonstrated to undergo large age-dependent changes, especially during early postnatal development (Fu *et al.*, 2011; Li *et al.*, 2017; Rouser *et al.*, 1968).

Metabolite and lipid measurements

Metabolomics and lipidomics are relatively new directions in the “omics” field. Metabolic processes are believed to be strongly dependent on the environmental factors, which results in technical difficulties or at least the need to choose between several options of metabolic measurement designs. By contrast, lipids are more stable in terms of environmental and technical variability. Furthermore, the recent technological progress in mass spectrometry analysis has allowed the rapid evolution of metabolite and lipid biomedical studies (Yang *et al.*, 2016).

The most commonly used methods for metabolite and lipid detection are liquid and gas chromatography (LC and GC) coupled with mass spectrometry (MS). Generally, the chromatographic separation of non-polar compounds is commonly conducted using LC to ensure greater coverage, whereas compact polar compounds are separated using GC technique. Before a GC run, polar groups in the compounds can be chemically derivatized to convert them into non-polar ones in order to make the molecules more

volatile (Dasgupta *et al.*, 2014). This changes the molecule's mass and structure, thus influencing further data analysis.

Due to chemical differences, the molecules of the sample are differentially retained in the chromatographic column and eluted at the different times, resulting in retention time differences, which allows separate downstream molecule capture and detection. After being eluted from the column, the molecule is ionized and the charged form or the ion fragments are detected in the mass spectrometer as mass-to-charge ratio (m/z). Mass-to-charge ratio, taken together with retention time, could be used for the compound differentiation in the data analysis steps (Gohlke, 1959; Yang *et al.*, 2016).

Depending on the experimental aim, one of two approaches in terms of analytical coverage can be chosen: targeted or global/untargeted analysis. Targeted analysis is powerful for studying defined sets of chemical compounds and specific metabolites. Untargeted analysis increases the coverage of detected molecules and allows the mapping of observed abundance differences to metabolic pathways (Hyötyläinen *et al.*, 2014; Yang *et al.*, 2009). Since the untargeted approach could be used to reconstruct the entire molecular phenotype, it is the preferred method for metabolome and lipidome analysis.

Modern human population diversity

Population variation is driven by natural selection under the influence of environmental factors. It provides population adaptation and underlies phenotypically beneficial traits. The Human Genome Project demonstrated that human individuals can be clustered into populations based on genetic data (Rosenberg, 2002; Turakulov *et al.*, 2003). One of the methodological approaches is to evaluate single nucleotide

polymorphisms (SNPs) of the whole or part of the genome and determine the loci that differentiate inter-population differences. It was confirmed that only 10-15% of genetic differences explain between-population variation and 85–90% of differences occur within a population (Romualdi *et al.*, 2002; Rosenberg, 2002). The overall genetic divergence, both at the SNP and copy number variation (CNV) levels, exposes the greatest variation within Sub-Saharan Africa with the rest of the human populations showing gradual out-of-Africa divergence (White *et al.*, 2003). Among this neutral variation, which reflects the migration history of the contemporary human populations (Nielsen *et al.*, 2017), there are several well-known examples of adaptive genetic differences, specific to particular populations or population groups. A typical example of positive selection in a population is a FY*O allele at the Duffy locus ensuring a resistance to malaria (Livingstone, 1984), which is fixed in sub-Saharan Africa but rare elsewhere in the world.

A similar example is the rs3827760 SNP: East Asians and native Americans unlike Europeans and Africans, carry a non-synonymous SNP in the coding region of the EDAR gene (Sabeti *et al.*, 2007), which is involved in hair, teeth and exocrine gland development (Botchkarev *et al.*, 2005). Experiments on mice models showed that Asian amino acid substitution resulted in thicker hair and larger salivary glands (Shie Hong Chang *et al.*, 2009).

In addition to genetic variation, differences among human populations have been shown at the levels of DNA methylation (Fraser *et al.*, 2012), in the expression of protein-coding genes (Hughes *et al.*, 2015), and expression of small non-coding RNA (microRNA) (Xiangyun Chang *et al.*, 2014; Rawlings-Goss *et al.*, 2014). Most of the

differences in gene expression were investigated in lymphoblastoid cell lines (Lappalainen *et al.*, 2013; Stranger *et al.*, 2012), but also in the placenta (Hughes *et al.*, 2015) and whole blood (Saw *et al.*, 2017). One study assessed inter-population gene expression differences in the brain in the prefrontal cortex of five East Asians, five Western Europeans, and four African Americans (Khrameeva *et al.*, 2014). Despite a limited sample size, the study indicated the presence of population differences at both mRNA and lipid abundance levels.

This observation is intriguing, given the population differences were previously reported in neuroanatomical brain composition in cortical thickness, volume, and surface area in the multiple regions (Bai *et al.*, 2012; Tang *et al.*, 2018). Moreover, the population differences were reported for episodic and semantic memory and executive functioning (Early *et al.*, 2013; Masel *et al.*, 2009), as well as the probability of neurological disorders. Specifically, African Americans were reported to have a lower risk of Parkinson's disease (Hemming *et al.*, 2011; Wright Willis *et al.*, 2010) and amyotrophic lateral sclerosis (Gundogdu *et al.*, 2014), but might be more at risk for Alzheimer's disease (Chin *et al.*, 2011).

As previously discussed, differences in the abundance of polar metabolites in the prefrontal cortex have been linked to neuropsychiatric disorders – including schizophrenia and psychosis – and have also been associated with cognitive abilities, such as memory and orientation (Crabtree *et al.*, 2018, 2016; Goldberg *et al.*, 2001). Moreover, the lipid and metabolite composition of the human brain was demonstrated to undergo substantial age-dependent rearrangements, especially during early postnatal

development (Fu *et al.*, 2011; Li *et al.*, 2017; Rouser *et al.*, 1968). These observations suggest that lipidome and metabolome composition of the human brain is an important component of the brain functionality, as well as a potentially important element determining brain disorder susceptibility and treatment efficiency.

Introgression of ancient human population

Recent studies demonstrated that the origin and ancient admixture of *Homo sapiens* influences the genetic similarity and differences of modern populations. Fossil and archaeological discoveries coupled with genetic variation analysis have shown that the modern *Homo sapiens* species emerged in Africa around 120 000 years ago, left Africa about 70 000 years ago and cohabitated with contemporaneous hominin species, such as *Homo erectus* and *Homo neanderthalensis*. Anatomically modern humans spread across the entire globe, reaching Australia approximately 40 000 years ago and the Americas approximately 16 000 years ago.

Modern humans differ dramatically from their closest evolutionary relatives in a number of ways. Most strikingly, they have developed rapidly changing and complex cultures that have allowed them to become much more numerous than any other primates and hominins and to spread to almost all parts of the planet. This unique historical development has at least to some extent biological roots. Although the expansion covered a variety of environments, the genetic variation is higher in African populations compared to the variation in non-Africans (Tishkoff *et al.*, 2002). This might be due to a small number of individuals leaving Africa to populate the rest of the globe, carrying only

a fraction of the genetic variation present in African individuals (Lohmueller *et al.*, 2008).

However, although large numbers of traits have been identified as being unique to humans or been suggested to be so (Tomasello, 2019; Varki, 2005), it has proven difficult to identify the genetic and biological underpinnings of such traits. Recent studies demonstrated the presence of Neanderthal genetic material in the genomes of all modern humans living outside of sub-Saharan Africa. Neanderthal introgression within the modern non-African human genomes varies from 1.5 to 4 percent in each individual (Sankararaman *et al.*, 2014). The admixture with Neanderthals was shown to bring a higher risk of type-II diabetes in South Asians and native Americans (Williams *et al.*, 2014). Moreover, the population origin of an individual was shown to be related to the incidence of gestational diabetes (Makgoba *et al.*, 2012).

The number of DNA mutations that distinguish modern humans from Neanderthals is relatively small: only 96 amino acid substitutions in 87 proteins (Prüfer *et al.*, 2014). Functional annotation of these 87 genes is incomplete and does not reveal the specific enrichment. To date, none of these amino acid substitutions have been linked to a function.

Thus, detailed study of each protein carrying human-specific substitutions may shed the light on the evolution of modern humans. For example, if we consider that brain development played a crucial role in the evolution of modern humans, the VCAM1 gene is involved in the maintenance of neural stem cells in the adult subventricular zone (Kokovay *et al.*, 2012) and five genes are expressed in the ventricular zone (CASC5,

KIF18A, TKTL1, SPAG5, VCAM1) and are associated with the mitotic spindle and the kinetochore (Prüfer *et al.*, 2014), which might have a functional role in neurogenesis (Fietz *et al.*, 2011).

One of the amino acid substitutions separating modern humans from Neanderthals occurred in the adenylosuccinate lyase (ADSL) protein. ADSL is a homotetrameric complex where three monomers contribute to each active site. The gene encoding ADSL in humans is located on chromosome 22q13.1–13.2 (Fon *et al.*, 1993). ADSL cleaves adenylosuccinate (S-AMP) into adenosine monophosphate (AMP) and fumarate. It further cleaves succinylaminoimidazole carboxamide ribotide (SAICAR) into aminoimidazole carboxamide ribotide (AICAR) and fumarate (Marie *et al.*, 2002). The compounds AMP, IMP, and GMP that occur at lower levels in the prefrontal cortex of humans than in chimpanzees and macaques are situated downstream of ADSL in the purine biosynthesis pathway, suggesting that a change in ADSL sequence could be the cause of reduced tissue concentrations of these compounds.

The amino acid substitution in ADSL results in an alanine to valine substitution at position 429 in the protein (A429V) and is present in almost all present-day humans whereas Neanderthals, Denisovans and all other primates, as well as most mammals, carry an alanine residue at position 429. The substitution is located in a protein domain that forms part of the substrate channel over the active site of the enzyme (Ray *et al.*, 2012) (Figure 1). Amino acid substitutions close to these positions causes lowered enzymatic activity and/or lower enzyme stability resulting in adenylosuccinate lyase deficiency in humans (Marie *et al.*, 2002). This condition is characterized by symptoms

that include psychomotor retardation, autism, epilepsy (Jurecka *et al.*, 2015; Nassogne *et al.*, 2000) and alterations in brain structures as observed with magnetic resonance imaging (Jurecka *et al.*, 2012; Zulfiqar *et al.*, 2013). The A429V substitution is therefore an attractive candidate for having functional consequences for metabolic variations in modern humans.

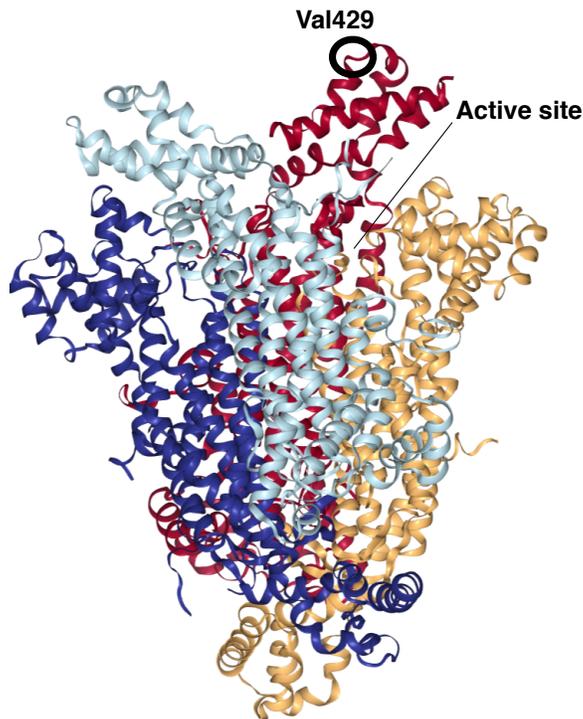


Figure 1. Molecular structure of human tetramer colored by chain. The position of the Val429 residue which distinguishes human ADSL from Neanderthal-like ADSL is indicated in one subunit by a dark circle.

Given the ability to detect such amino acid substitutions and other genetic differences between humans described above, we aim to investigate the differences between three contemporary human populations and Neanderthals. Using mass-

spectrometry approach, we can determine what, if any, variations exist in the brain metabolomes and lipidomes of these populations.

Chapter 3. Results

Differences in metabolome organization in modern populations

Lipidome and metabolome variation analysis

We assessed the abundance of lipids and polar metabolites in the prefrontal cortex samples of 146 Han Chinese (HC), 97 Western European (WE), and 60 African American (AA) individuals. For each population, the ages of individuals covered most of the lifespan: from birth to 71 years of age (Figure 2A). For each individual, the lipids and metabolites were extracted from the same cortical gray matter sample dissected from the dorsolateral region of the prefrontal cortex. Tissue preservation of the samples was assessed using postmortem interval duration (PMI) and RNA integrity number (RIN) measured for a subset of individuals.

The lipid abundance measurements were conducted using liquid chromatography coupled with untargeted mass spectrometry (LC-MS) in positive and negative ionization modes. The LC-MS measurements yielded a total of 1,670 distinct lipid peaks not affected by the confounding factors, such as extraction batch, mass spectrometry loading order, and PMI. Among them, 900 peaks were computationally annotated based on mass-to-charge ratio values using LIPID MAPS Structure Database (LMSD) (Sud *et al.*, 2007). All analyses of the lipidome population differences were based on the intensities of 1,670 detected lipids unless indicated otherwise.

The metabolite abundance measurements were conducted using gas chromatography coupled with mass spectrometry (GC-MS). The GC-MS measurements

yielded 258 confounder-free compounds identified and annotated using previously analyzed metabolite standards. All analyses of the metabolome population differences were based on the intensities of these 258 metabolites unless indicated otherwise.

The overall lipidome variation analysis conducted using t-distributed stochastic neighbor embedding (t-SNE) based on the abundance of 1,670 detected or 900 computationally annotated lipids showed strong separation of the youngest individuals from the rest. The same result was observed based on the abundance of 258 detected polar metabolites (Figure 2C). Nonetheless, the trend distinguishing samples of different populations was also apparent in both the lipid and metabolite data. Correspondingly, the variation analysis of the lipid abundance levels indicated that age explained 28% of the total variation and population identity 3%. Other factors, such as sex, RIN, and PMI each explained less than 1% of the total lipidome variation. Similarly, for the polar compounds, population identity explained 6% of the total variation, and the other factors less than 2.5% each.

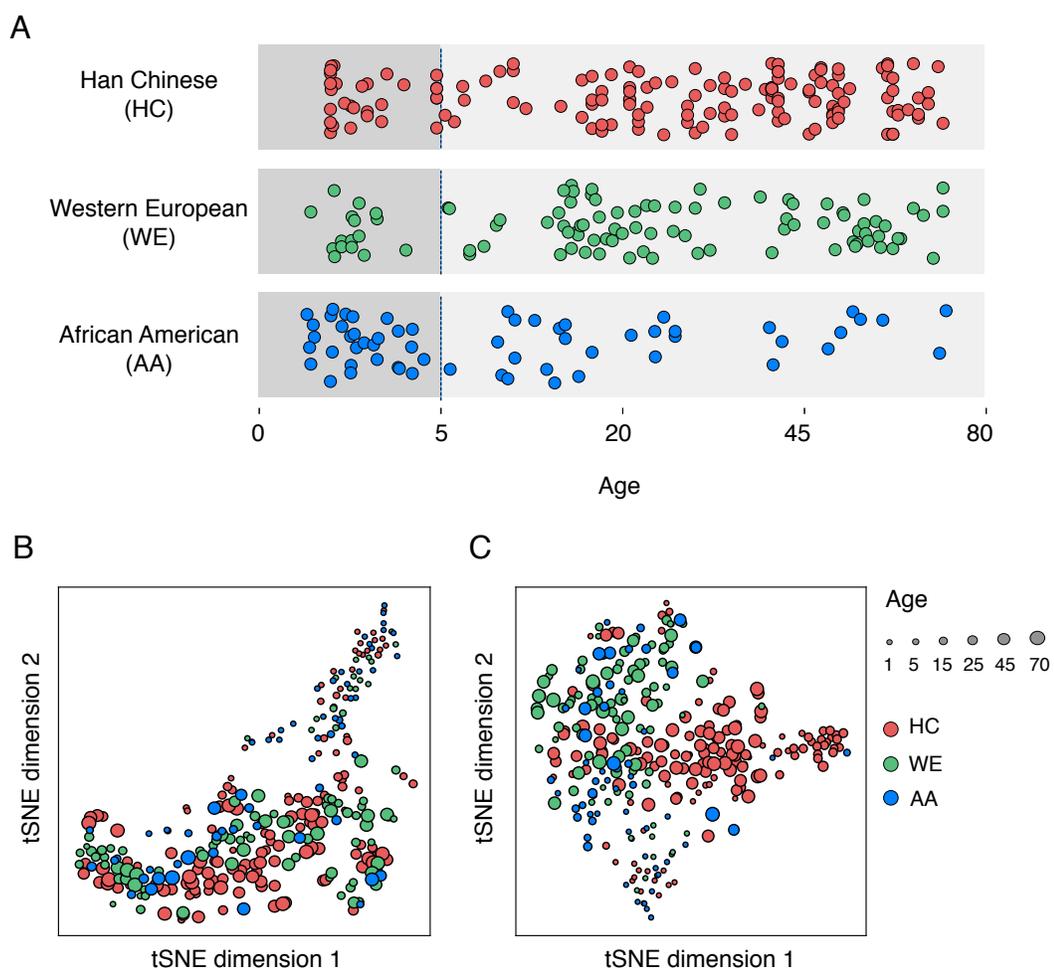


Figure 2. Assessment of the lipid and metabolite abundance variation.

(A) Age distribution of samples. Each circle represents an individual. The circle colors correspond to populations: red – Han Chinese (HC), green – Western Europeans (WE), blue – African Americans (AA). The background color delineates two datasets: the darker shade – samples with ages less than 5 years (DS:0-4) and the lighter shade – samples with ages greater than 5 years (DS:5-71). Lipid (B) and metabolite (C) abundance variation among individuals visualized using t-distributed stochastic neighbor embedding (t-SNE). Each circle represents an individual. The sizes of the circles are related to the individuals' age – larger circles represent older individuals. Colors represent populations, as described above.

Statistical analysis of lipid and metabolite differences among populations

Consistent with previous reports (Li *et al.*, 2017), variation analysis indicated the presence of strong age-dependent lipidome and metabolome differences between samples of younger ages and the remaining individuals (Figures 2B, 2C). To reduce the influence of age on the inter-population variation, we separated the samples into two datasets: DS:0-4 ($n = 74$, ages less than five years), and DS:5-71 ($n = 229$, ages from five to 71 years) (Figure 2A). We then searched for lipid abundance differences characteristic of each population by comparing it to the other two populations. To equalize the statistical power, we subsampled the same number of individuals per population in every comparison 100 times ($n = 13$ for DS:0-4 and $n = 25$ for DS:5-71) and identified lipids showing population-specific abundance levels in each subsampling. In DS:0-4, this analysis revealed no lipid abundance differences specific to any particular population, with a marginally higher number of differences specific to WE (median = 1 for WE, median = 0 for HC and AA, t-test, Benjamini-Hochberg corrected $p < 0.05$; Figure 3A). By contrast, in DS:5-71, HC population differed from the other two by the median of 90 lipids, while no lipids showed abundance levels specific to WE and AA in the average of 100 sample subsets (t-test, Benjamini-Hochberg corrected $p < 0.05$; Figure 3A). The specific lipidome behavior of HC population in DS:5-71 was not due to the difference in statistical power between DS:5-71 and DS:0-4, as shown by subsampling the same number of individuals in both datasets (Figure 3A). Restricting the analysis to well-preserved DS:5-71 samples defined based on RNA conservation and an empirically

determined RNA quality threshold (Gallego Romero *et al.*, 2014) ($RIN > 7$, $n = 82$) retained an evident excess of the HC-specific lipid abundance differences, compared to AA and WE population-specific differences. Similarly, exclusion of lipids showing even weak correlation between the abundance and PMI duration (definition of PMI effect threshold relaxed to: absolute value of correlation coefficient $r = 0.137$, $p = 0.1$) did not alter the results.

Analysis based on all DS:5-71 samples (without subsampling) yielded 395 lipids showing abundance levels specific to HC population (t-test, Benjamini-Hochberg corrected $p < 0.05$). Notably, the comparison to an independently generated published adult cortical lipidome dataset consisting of five HC, five WE, and four AA individuals (Khrameeva *et al.*, 2014) confirmed the identified lipid concentration differences specific to HC population (Spearman correlation test, $p = 0.003$).

The same statistical analysis applied to the polar metabolite dataset produced similar results. While there were no metabolite concentration differences specific to either of the three populations in DS:0-4 the median of 93 metabolites showed abundance levels particular to HC, 23 – to WE, and 14 – to AA in DS:5-71 (t-test, Benjamini-Hochberg corrected $p < 0.05$; Figure 3B). Similar to the lipid data, the excess of metabolic differences particular to HC population in DS:5-71 remained robust after subsampling the same number of individuals in both DS:5-71 and DS:0-4 (Figure 3B). Restriction of analysis to well preserved DS:5-71 samples ($RIN > 7$, $n = 80$) or exclusion of polar metabolites showing even weak correlation between the abundance and PMI duration (definition of PMI effect threshold: absolute value of correlation coefficient $r =$

0.170, $p = 0.1$) retained an evident excess of the HC-specific metabolite abundance differences.

Analysis based on all DS:5-71 samples (without subsampling) yielded 166 HC-specific metabolite differences (t-test, Benjamini-Hochberg corrected $p < 0.05$)

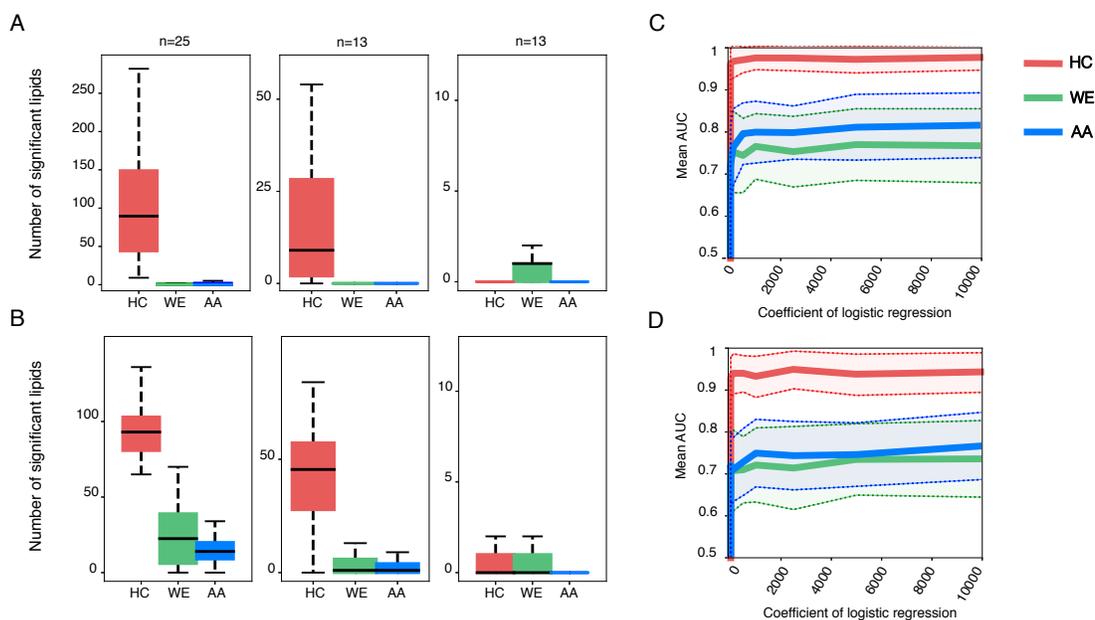


Figure 3. Lipidome and metabolome population-specific differences in the DS:0-4 and DS:5-71 datasets.

Number of lipids (A) and metabolites (B) with significant abundance differences between one population and the other two combined, estimated by subsampling n individuals from each of the three populations 100 times. The numbers of subsampled individuals n used in analysis are marked on top of the panels. The results are shown for DS:5-71 (left, $n = 25$), DS:5-71 with same number n as DS:0-4 (middle, $n = 13$), and DS:0-4 (right, $n = 13$). The colors represent populations: red – Han Chinese (HC), green – Western Europeans (WE), blue – African Americans (AA).

The mean Area Under the ROC Curve (AUC) estimates for the lasso logistic regression models separating samples from one population and samples from the other two combined calculated using different values of hyperparameter C (inverse of regularization strength). The models were based on the lipid (C) and metabolite (D) abundance in DS:5-71 samples. Lines correspond to the means of AUC values estimated on different test sets. The shaded areas indicate the standard deviations of AUC values estimated on different test sets. The colors correspond to populations, as described above.

Population classification using machine learning

To test whether the populations could be distinguished reliably based on lipid or polar metabolite abundance values, we classified samples using lasso logistic regression model. Because of the strong effect of age on compound abundances resulting in the separation of samples from very young individuals (Figures 2B, 2C), the classification procedure was applied to DS:5-71 only and not to the complete set of samples. The classification procedure was not applied to the DS:0-4 separately because of insufficient sample size.

For the lipid data, the resulting model accurately separated the HC population from the other two (area under the curve AUC = 0.97) (Figure 3C). The separation of AA population, as well as WE population, was notably less accurate (AUC = 0.8 and 0.76, respectively), although still significantly better than expected by chance (Figure 3C). We used stability selection procedure (Meinshausen *et al.*, 2008) to define lipid predictors of HC population. These predictors ($n = 200$) overlapped well with statistically defined lipid abundance differences (hypergeometric test, $p = 3.5 \times 10^{-54}$). Notably, the abundance

differences of lipid predictors between HC and the other two populations correlated with the differences determined using an independent published dataset (Khrameeva *et al.*, 2014) (Spearman correlation test, $p = 0.005$). Furthermore, validation of the model on this independently generated published dataset resulted in good classification of HC individuals (AUC = 0.89). It has to be noted that model accuracy estimates were limited in this case by the size of the published dataset.

Application of the logistic regression model to metabolite DS:5-71 dataset similarly resulted in significantly higher classification accuracy for HC individuals: AUC = 0.94 compared to AUC = 0.72 for WE, and AUC = 0.74 for AA (Figure 3D). Stability selection procedure yielded 50 metabolite predictors of HC population, which overlapped well with the statistically defined differences (hypergeometric test, $p = 0.002$).

Age dynamics of population difference

The statistical analysis of lipid and metabolite abundance yielded detectable HC-specific differences only in DS:5-71 samples. Similarly, the accuracy of the logistic regression model trained on DS:5-71 data remained mostly unchanged when applied to classification of AA and WE samples with ages from 0 to 4 years, but dropped drastically when applied to classification of HC samples from 0-4 years age interval. Notably, this drop in accuracy was detected at approximately two years of age, even though all samples with age < 5 years were excluded from the training set of the logistic regression model. These results suggest that the detected HC-specific differences cannot be generalized to HC individuals younger than two years of age.

To assess the relationship between the population differences and individuals' age, we divided all individuals into six age groups A1-A6, separated at 1, 5, 15, 25, 45 years of age and containing 24-75 individuals each (Figure 4A). We then subsampled the same number of individuals of each population within each age group ($n = 4$) 1,000 times and identified top 50 lipids or top 20 metabolites showing the most consistent abundance differences between each population pair in each age group (t-test, nominal $p < 0.1$). We used the union of these lipids or metabolites to construct a set of age-unbiased population-distinguishing compounds. The distances calculated from correlations of population-mean abundances of these compounds showed strong age-dependent behavior of HC-related differences for both lipids and metabolites (Figures 4B, 4C). Specifically, the distances between HC and the other two populations increased substantially after the first year of life and then decreased after 20-30 years of age (Figures 4B, 4C). Thus, age groups contributed unevenly to the separation of HC samples from the other two populations in the DS:5-71 sample set, with the strongest contribution provided by the young adult groups. This pattern was even more pronounced for HC-specific compound abundances. By contrast, AA-WE distance did not show any substantial increase along the lifespan (Figures 4B, 4C). Additionally, among all six age groups, the HC-AA and HC-WE distances were the smallest in A1 age group. This result is consistent with inaccurate logistic regression model performance for classification of HC individuals younger than two years of age.

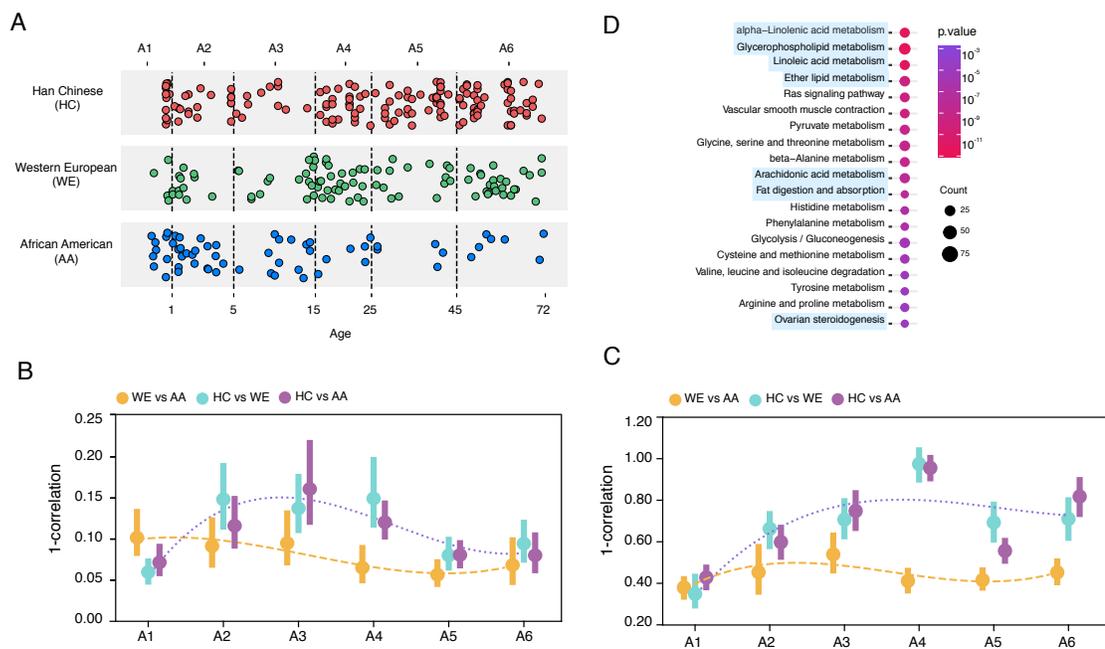


Figure 4. The age-dependent dynamics of lipidome and metabolome differences between populations.

(A) Sample distribution across six age groups. Each circle represents an individual. The colors represent populations: red – Han Chinese (HC), green – Western Europeans (WE), blue – African Americans (AA). The x-axis labels indicate the age groups’ boundaries in years.

Pairwise population differences estimation based on the abundance of age-unbiased population-distinguishing lipids (B) and metabolites (C). Differences were calculated in each age group A1-A6 using correlations of population-mean abundances based on four samples subsampled from each population. Y-axis represents distance values calculated as one minus these correlation values. Circles represent the median distance values estimated by subsampling within each age group 10,000 times. Vertical lines extend to the upper and lower quartile values in each age group. The dotted blue line represents a smooth spline curve fitted to the average of the HC-WE and HC-AA distances. The dashed orange line represents a smooth spline curve fitted to the WE-

AA distance. (D) Pathway enrichment analysis. Shown are the top 19 pathways (Benjamini-Hochberg corrected hypergeometric test $p < 0.001$) that show enrichment of genes linked to HC-specific lipids and metabolites. HC-specific lipids and metabolites defined from stability selection were used in this analysis. Circle sizes represent the number of genes linked to HC-specific lipid and metabolite compounds. Circle colors correspond to Benjamini-Hochberg corrected hypergeometric test p-values. Pathways associated with lipid metabolism are shaded in light blue.

Functional characterization of HC-specific lipid and metabolite differences

We assessed potential functions of lipids and polar metabolites distinguishing adult HC individuals by testing their enrichment in functional pathways defined by KEGG (Kyoto Encyclopedia of Genes and Genomes) (Kanehisa *et al.*, 2017). The enrichment analysis involved 900 computationally annotated lipids and 258 polar annotated metabolites and was based on the comparison between genes linked to HC-specific compounds and genes linked to the other detected compounds according to KEGG database. The analysis yielded a total of 35 significantly enriched pathways, including ten pathways associated with amino acid metabolism and seven pathways associated with lipid metabolism (hypergeometric test, Benjamini-Hochberg corrected $p < 0.05$). Notably, all seven pathways associated with lipid metabolism were present in the top 19 enriched terms (hypergeometric test, Benjamini-Hochberg corrected $p < 0.001$; Figure 4D).

Differences in metabolome organization in ancient populations

Metabolic changes unique to humans

Due to absence of samples for metabolic experiments, the task to evaluate lipids and metabolites of ancient humans is challenging. We firstly decided to understand the ancient metabolome fingerprint and identified metabolites that have changed their concentration in humans relative to monkeys and apes. We measured compound concentrations in prefrontal cortex, primary visual cortex, cerebellum, skeletal muscle and kidney in four humans, four chimpanzees, and four macaques using mass spectrometry coupled with capillary electrophoresis (CE-MS), a technique suitable for detection of small hydrophilic compounds. The number of metabolites annotated in the five tissues varied between 166 and 197, 160 and 209, and 145 and 192 in the three species, respectively (Figure 5A,B).

In each tissue, we identified metabolites that did not differ significantly in concentrations between the macaques and the chimpanzees, but differed significantly, and in the same direction, between humans and macaques as well as between humans and chimpanzees. In skeletal muscle and kidney we find no such metabolites. In contrast, in two of the three parts of the brain analyzed we find metabolites that differ in their concentration in humans relative to the other two primates.

In cerebellum, 22 metabolites have higher concentrations in humans whereas no metabolites have lower concentrations. Eighteen of the 22 metabolites are amino acids. In prefrontal cortex we detected five metabolites with lower concentrations in humans and none with higher concentrations (BH-corrected, $p < 0.05$). Three of the five metabolites

are purines (inosine monophosphate (IMP), guanosine monophosphate (GMP), adenosine monophosphate (AMP)) and the other two are NAD⁺ and UDP-N-acetylglucosamine.

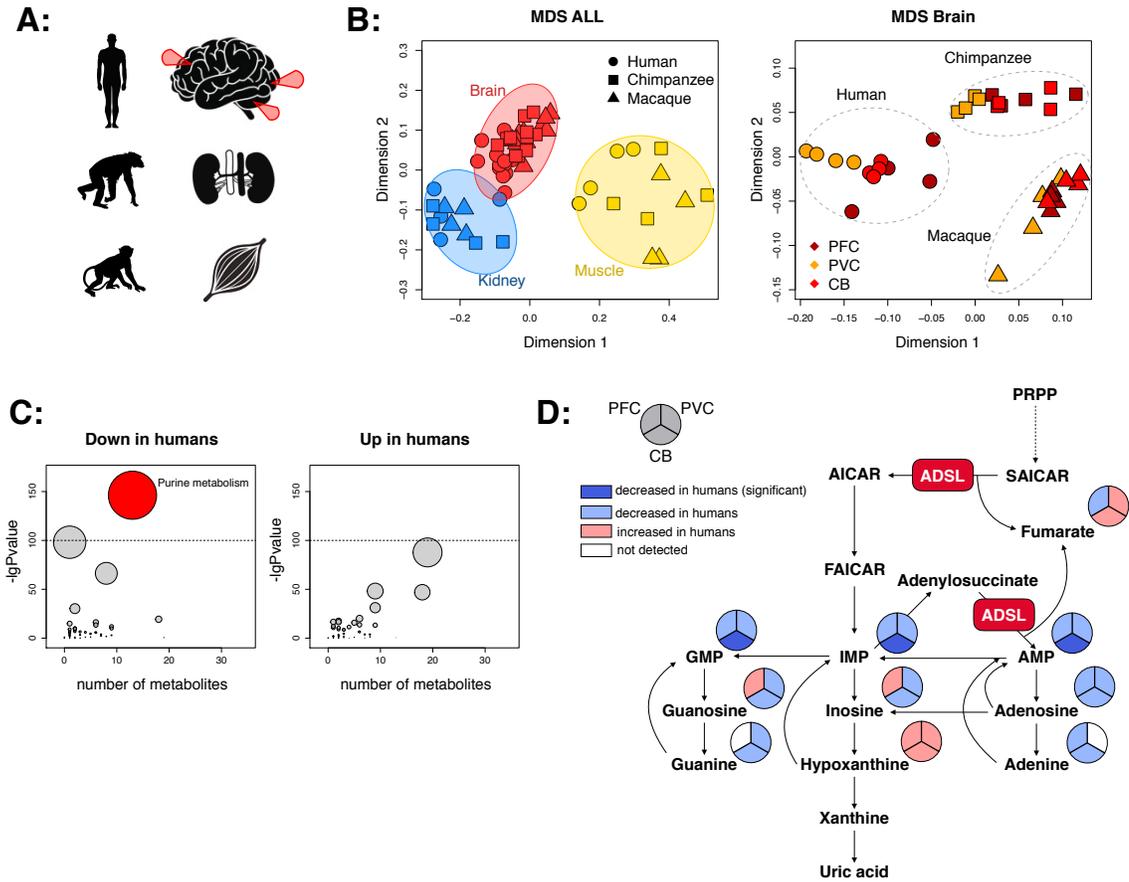


Figure 5. CE/MS metabolomic data of three primate species.

(A) Data overview. Metabolite concentrations were measured in the prefrontal cortex (PFC), primary visual cortex (PVC), cerebellum (CB), kidney and muscle in three primate species, 4 individuals of each: rhesus macaque, chimpanzee and human.

(B) The relationship among all tissue samples and among the brain tissue samples plotted in two dimensions using a multidimensional scaling algorithm (MDS). MDS plots show Euclidean

distances based on concentrations of detected metabolites. Symbols represent species, colors represent tissues, and points represent individual samples.

(C) Enrichment of pathways associated with metabolites downregulated and upregulated in human. Plots show the cumulative effect calculated as $-\log_{10}$ Fisher's test p-value for pathways across tissues (y-axis) and the number of detected metabolites in each pathway (x-axis). Size of the circle is proportional to $-\log_{10}$ p-value.

(D) The simplified schematic representation of *de-novo* purine biosynthesis pathway showing human-specific changes in three brain regions.

Metabolic pathways unique to humans

To identify metabolic pathways that may be more or less active in humans than in other primates, we linked metabolites with higher or lower concentrations in humans compared to both chimpanzee and macaque to genes and metabolic pathways using the KEGG database.

Between 8 and 12 pathways are associated with higher metabolite concentrations in the five human tissues. Out of eight pathways identified in the prefrontal cortex, seven involve amino acid metabolism, and seven and eight out of nine pathways involve amino acids in the visual cortex and the cerebellum, respectively. In muscle and kidneys, seven and five of the eleven and 12 pathways identified, respectively, also involve amino acid metabolism. Thus, several aspects of amino acid metabolism seem to be increased in humans relative to other primates.

Between four and six pathways are associated with lower metabolite concentrations in the five tissues. Amino acid and peptide metabolism make up one to three of these

pathways in the different tissues, suggesting that amino acid metabolism has changed in several ways in humans relative to the other primates, resulting in that metabolites are present in increased as well as decreased concentrations.

In contrast, metabolites in two pathways are consistently present at lower rather than higher concentrations in humans. One of these pathways is oxidative phosphorylation, which is between the two top pathways that show decreased metabolite concentrations in all three brain regions analyzed. Because oxidative phosphorylation is not affected in muscle and kidneys it seems that the human brain differs from ape brains in that oxidative phosphorylation in mitochondria is less active.

The second pathway that stands out in all five tissues is purine biosynthesis where a number of metabolites are present at lower levels (Figure 5C). Thus, purine biosynthesis is decreased in humans relative to apes in the brain as well as in other organs. If we focus on a more narrow definition of the purine biosynthesis pathways (Marie *et al.*, 2004) (Figure 5D), a significant decrease in purine biosynthesis is seen in the three brain regions and not in muscle and kidney. Furthermore, three of the five metabolites, which individually show significant human-specific concentration decreases in prefrontal cortex, IMP, GMP and AMP, are end products of purine *de novo* biosynthesis (Figure 5D). Thus, purine biosynthesis stands out as down-regulated in humans, particularly in the brain.

The metabolome of the humanized Adsl mice

To investigate if the A429V in ADSL substitution may be involved in the reduced purine biosynthesis seen in humans we introduced a nucleotide substitution resulting in

this amino acid substitution into the *Adsl* gene of a mouse. Adjacent to the amino acid of interest, at position 428, rodents carry an arginine residue, whereas other mammals including primates carry glutamine. To avoid possible effects of the rodent-specific arginine residue on the function of the neighboring amino acid at position 429 we in addition introduced a nucleotide substitution resulting in an R428Q substitution in the *Adsl* gene of a C57BL/6 mouse by homologous recombination. The two mutations segregate in Mendelian ratios in the mice. Animals heterozygous and homozygous for the two substitutions show no overt phenotypic difference to their wild type littermates.

We analyzed the metabolome of the nine tissues (prefrontal cortex, cerebellum, lung, liver, heart, kidney, muscle, spleen, testis) from 9-12 adult homozygous humanized and their wild-type littermates by GC-MS (Figure 6A). We similarly analyzed the eight tissues from 10-12 one-week-old pups. The number of metabolites detected varied between 176 and 273 in the adult mice and 310 and 347 in the young mice. Among those, an average of 176 (median = 273) were detected in at least 50% of the individuals in each of the nine tissues. A principle components analysis using the concentrations of all metabolites detected revealed one to two outlier samples per tissue. These were excluded from further analyses. Including these samples in the analyses did not qualitatively affect results.

Among the organs analyzed, only the brain showed significant differences in metabolite concentrations between the wild type and humanized mice. Specifically, 36 metabolites showed significant concentration differences in cerebellum of 12-weeks-old mice (permutations, $p < 0.05$) and 45 metabolites showed significant differences in the

cerebral cortex of one-week-old mice (permutations, $p < 0.05$). Thus, the metabolic effects of the two mutations introduced in ADSL are particularly pronounced in the central nervous system.

The concentration differences detected in cerebellum of the 12-weeks-old mice correlated with differences observed in cerebral cortex, even though differences in cortex did not pass the significance cut-off in our permutation test (Pearson correlation, $r = 0.84$, $p < 0.0001$, $n = 29$). Similarly, concentration differences detected in cerebral cortex in one-week-old mice correlated with differences observed in cerebellum in the same mice (Pearson correlation, $r = 0.77$, $p < 0.0001$, $n = 45$). By contrast, the correlation between metabolic differences detected in the brain and the other tissues was weaker (Figure 6B). The concentration differences between wild type and humanized mice furthermore correlated between the one-week-old and 12-week-old mice both in cortex and in cerebellum (Pearson correlation, $r = 0.49$ and $r = 0.52$, $p < 0.005$, $n = 28$ and $n = 31$, respectively). Thus, in the humanized mouse, the effects of the substitutions in ADSL are seen in the cerebral cortex and cerebellum in both young and adult animals.

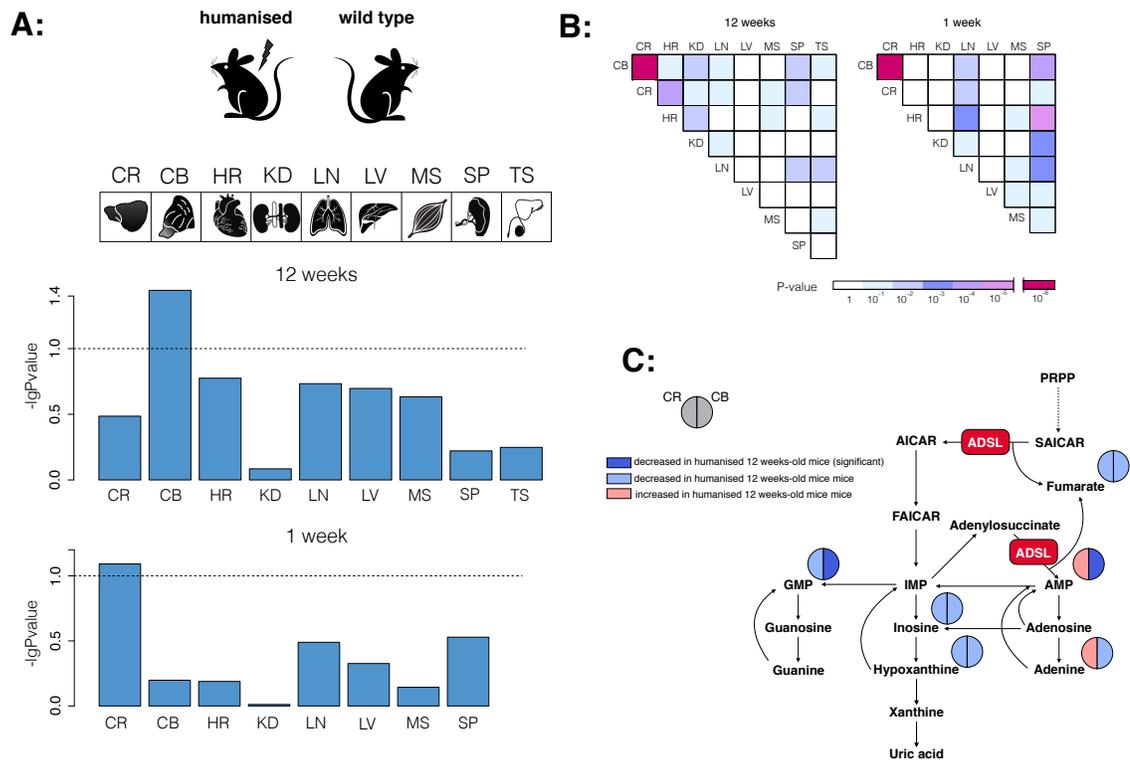


Figure 6. GC/MS metabolomic data of humanized and wild-type mice.

(A) Data overview. Metabolite concentrations were measured in the cortex (CR), cerebellum (CB), heart (HR), kidney (KD), lung (LN), liver (LV), muscle (MS), spline (SP) and testis (TS) in humanized and wild-type mice. Barplots show $-\log_{10}$ p-values of metabolite concentration differences between humanized and wild-type samples (Student's test, 1000 permutations) for 1 week- and 12 week-old mice.

(B) Correlations between cerebral cortex and other tissues in 1 week- and 12 week-old mice. Colors show Spearman correlation p-values for a subset of metabolites with significant metabolite concentration differences between humanized and wild-type samples at the Student's test corrected p-value cutoff equal to 0.05.

(C) The simplified schematic representation of *de-novo* purine biosynthesis pathway showing metabolite concentration changes specific for humanized mice, in two brain regions.

Purine biosynthesis in the humanized mice

In the 12-weeks-old mice, we detected six metabolites within the purine biosynthesis pathway. In the humanized mice, four of these metabolites had lower concentrations than in the wild-type mice in the cerebral cortex and all of them had lower concentrations in the cerebellum (binomial test for cortex and cerebellum, $p = 0.04$). In the one-week-old mice, we detected 14 metabolites in the purine biosynthesis pathway. In the humanized mice, 8 of these had lower concentration in the cerebral cortex and in the cerebellum when compared to their wild-type littermates.

The six metabolites detected in 12-week-old mice were also detected in the primate brains. Four of these had lower concentrations in the human brain, including AMP and GMP in prefrontal cortex, the visual cortex, and in cerebellum (binomial test, $p = 0.04$). Furthermore, four of the six metabolites with lower concentrations in cerebellum in the humanized mice had lower concentrations in human cerebellum compared to other primates: AMP, GMP, inosine, and adenine (Figure 6C). Thus, several changes in concentrations of compounds in purine biosynthesis seen in the humanized mice recapitulate differences seen when the metabolomes of the human brains are compared to the brains of chimpanzees and macaques.

Activity and stability of modern humanized mouse ADSL

To investigate how the humanized form of the mouse ADSL enzyme may influence purine biosynthesis, we synthesized mouse wild-type (wt) ADSL and mouse A429V ADSL and inserted them in expression vectors that include N-terminal polyhistidine tags

(Lee *et al.*, 2007). To analyze the enzymatic activities of the two forms of the enzyme we tested each of the two activities of ADSL: the conversion of SAICAR to AICAR, and the conversion of SAMP to AMP, in the presence of excess substrate by measuring the rate of production of AICAR and AMP. We found no differences in the kinetics of either reaction between wt and A429V ADSL (t-test, $p > 0.05$).

We next tested if the A429V substitution influences the stability of the secondary structure of the enzyme by measuring the circular dichroism spectra of the purified proteins at 222 nm while heating them from 55°C to 80°C at a rate of one degree per minute. We next investigated if the R428Q substitution that was introduced into the humanized mice together with the A429V substitution affects the conformational stability of the protein. To do this, we generated expression vectors that carry wild-type murine ADSL, ADSL with the A429V substitution, ADSL with only the R428Q substitution, and ADSL with the A429V substitution and the adjacent R428Q substitution. We then compared the secondary structure stability of all four protein variants. Mouse ADSL protein with alanine at position 429 was more stable than mouse ADSL protein with valine at position 429. The effect of the A429V substitution was not influenced by the presence or absence of the adjacent R428Q substitution at position 428.

Thus, the A429V substitution does not affect the kinetics of the murine ADSL enzyme. However, it destabilizes the secondary structure of the protein as it was shown previously *in vitro* (Van Laer *et al.*, 2018).

Validation of metabolic changes using cell lines

To investigate how the A429V substitution affects the metabolome of living human cells, we used CRISPR-*Cas9* to introduce the nucleotide substitution in the *ADSL* gene in human pluripotent 409B2 cells that results in the reversion of the valine residue to the ancestral, Neanderthal-like alanine residue. We isolated three independent cell lines where we verified that the intended nucleotide substitution had occurred by sequencing a segment of the *ADSL* gene. We also isolated six independent lines that had been subjected to the editing procedure but did not exhibit any mutation in the sequenced DNA segment. We expanded 10 separate cultures of the three edited cell lines and 19 cultures of the control lines where position 429 in the inferred protein sequence had not been changed. Cell cultures metabolites were then analyzed by LC-MS in both positive and negative modes.

A total of 10,673 metabolites were detected. Among these were twelve metabolites from purine biosynthesis and three out of these (aminoimidazole ribonucleotide (AIR), hypoxanthine, guanine) were present in significantly lower amounts in the wild type than the ancestralized cells (t-test, $p < 0.01$). Strikingly, out of nine detected metabolites downstream of *ADSL*, all have lower concentrations in the wild type cells (binom. test, $p < 0.01$). Thus, in human cells, the ancestral version of *ADSL* supports a higher level of purine biosynthesis than the present-day, modern human version.

Comparison to chimpanzee cells

To compare these results to purine biosynthesis in chimpanzees, we similarly analyzed nine cell cultures from three different chimpanzee pluripotent cell lines. Out of

1,286 metabolites that differ significantly (t-test, $p < 0.01$) between the human and chimpanzee wild type cells, three (AIR, guanosine, guanine) are in purine biosynthesis. The same nine metabolites as in the comparison to the ancestralized cells were detected in the chimpanzee cells. Similar to the situation in the ancestralized human cells, they are present in lower concentrations in the human than in the chimpanzee cells (binom. test, $p < 0.01$). Thus, the reversal of the A429V substitution in ADSL in human cells results in an increase of purine biosynthesis similar to what is observed when chimpanzee cells are compared to human cells.

Purine biosynthesis in autistic metabolome

Since alteration in ADSL gene disrupts the function of the enzyme characterized by severe psychomotor delay and autistic features, we examined the purine biosynthesis pathway in ASD individuals. We compared metabolome of prefrontal cortex grey matter of 32 ASD patients (2-60 years old) and 40 control individuals (0-62 years old) and detected 1,366 metabolites not affected by experimental batch effects and post-mortem delay. Out of 1,366 metabolites 202 yielded significant concentration differences between ASD and control samples (ANCOVA, BH-corrected $p < 0.05$).

Hierarchical clustering of 202 ASD-related metabolites revealed specific age-dependent patterns clustered in the four modules. Genes linked to ASD-related metabolites using KEGG annotation were significantly overrepresented in a total of 16 pathways (hypergeometric test, BH-corrected $p < 0.05$). Remarkably, the two pathways containing enzymes showing the greatest autism associated differences in substrate binding energy, glutathione metabolism and purine metabolism, also showed the

strongest enrichment of metabolites showing concentration differences in autism, as well as genes linked to these metabolites. Metabolites involved in glutathione metabolism, including glutathione itself, were predominantly located in the module 3 and showed lower concentrations in autism compared to controls. Metabolites involved in purine metabolism were predominantly located in the module 2 characterized by lower concentrations in autism samples peaking in young adults. Moreover, module 4, which contained 52 metabolites, was enriched in four of the 16 pathways, including strong enrichment in purine and pyrimidine metabolism pathways.

In particular, we identified 14 metabolites from purine biosynthesis, seven of them show the decrease of concentrations in control unaffected by age, including AMP and GMP, that demonstrate consistent drop in human wild type cell lines and humanized mice (Figure 7). Additionally, two metabolites, adenine and guanine are decreased in controls of age range 10-25 years.

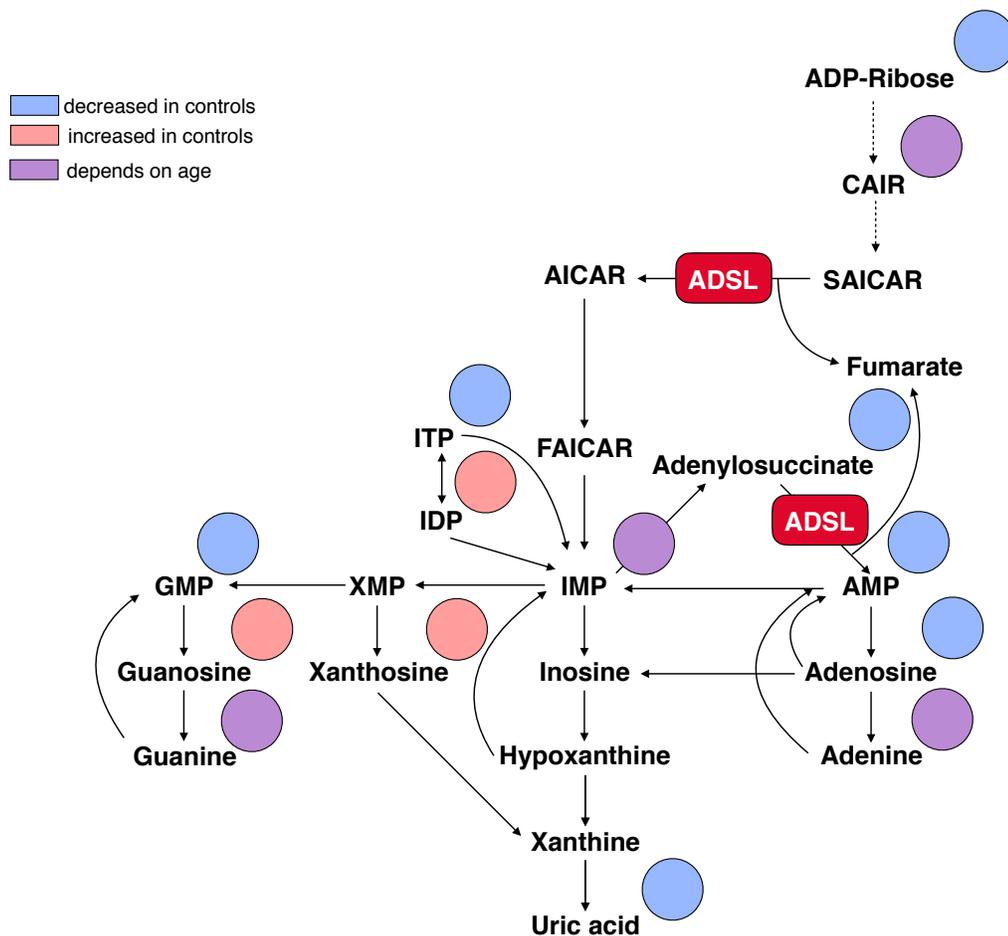


Figure 7. The simplified schematic representation of *de-novo* purine biosynthesis pathway showing metabolite concentration changes in control individuals compared to ASD individuals.

Evolution of purine biosynthesis metabolites and ASD

Similar to ancient population metabolome analysis we assessed the ASD metabolome placement on the evolutionary lineages. We measured metabolic abundances in prefrontal cortex of 40 chimpanzees (0-42 years old) together with human control and ASD samples in a random order. Experimental and computational procedures were the same as for the human samples, and resulted in 1,366 highly confident metabolites. MDS

analysis based on these metabolites revealed that the samples segregated predominantly according to age and species (Figure 8).

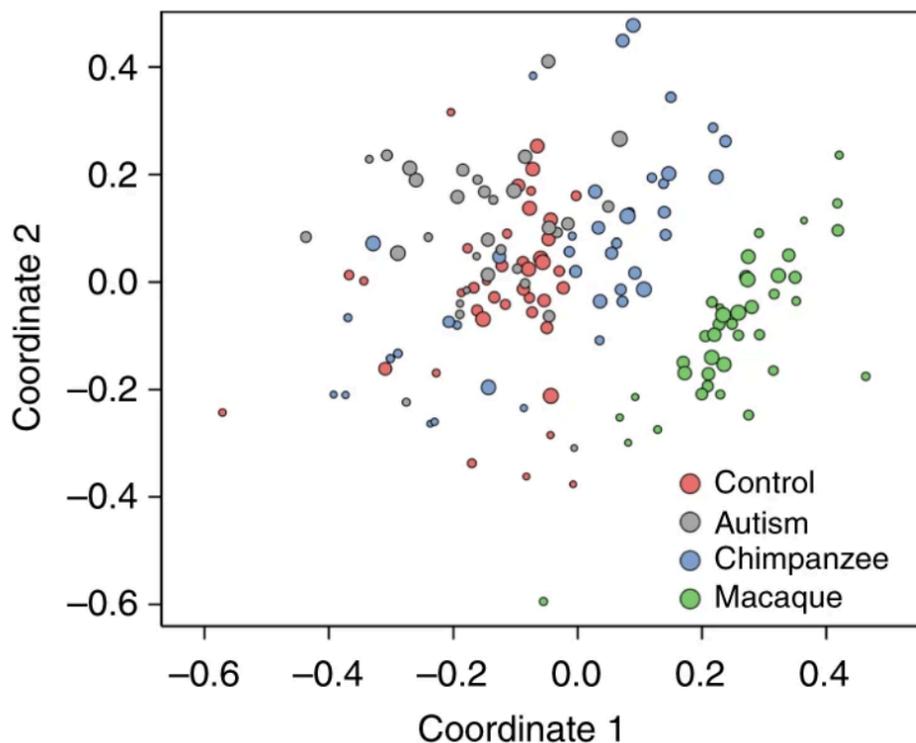


Figure 8. Metabolic similarity of ASD, unaffected controls, chimpanzees, and macaques. Each circle represents an individual (grey: ASD; red: controls; blue: chimpanzees; green: macaques). The size of the circles is proportional to the individual's age (smaller circles correspond to younger individuals) (Kurochkin *et al.*, 2019).

The identification of human-specific changes found 170 metabolites and human-specific concentration differences agreed well with the differences calculated using the published metabolome dataset (Khrameeva *et al.*, 2014) (Pearson correlation, $r = 0.71$, $p < 0.01$). Genes linked to the metabolites were significantly overrepresented in 27 KEGG pathways (hypergeometric test, BH-corrected $p < 0.05$). In accordance with the results

demonstrated above, these pathways include purine metabolism and overlap significantly with pathways enriched in ASD-related metabolic differences (Fisher test, $p < 0.01$). Herewith, the human-specific metabolites distribution between the four ASD modules was not uniform. Module 1 contained fewer human-specific metabolite intensity differences compared with the average, while module 4 contained approximately five times more (Wilcoxon test, $p < 0.01$), additionally enriched in purine and pyrimidine metabolism. Moreover, the abundances of adenylosuccinate, AMP, GMP and adenosine share the same pattern, having concentrations in ASD samples as an intermediate between unaffected controls and chimpanzee samples.

Chapter 4. Discussion

While lipids constitute the majority of the human brain's organic material and are essential for brain functionality, only a handful of studies to date examined human brain lipidome composition (Hunter *et al.*, 2018; Khrameeva *et al.*, 2014; Li *et al.*, 2017; Rouser *et al.*, 1968). Among them, one study assessed lipid composition of the prefrontal cortex in 14 individuals representing three populations, Han Chinese (HC), Western European (WE), and African American (AA), suggesting possible lipid abundance differences among populations (Khrameeva *et al.*, 2014). Our study expanded this work by including 303 individuals of different ages representing the same three populations. Our analysis indicates the robust presence of lipid and polar metabolite abundance differences distinguishing the prefrontal cortex composition of Han Chinese (HC) individuals from that of Western Europeans (WE) and African Americans (AA). The observed difference between HC individuals and the other two populations is age-dependent: it peaks at approximately 20 years of age and is absent during the first year of life. The difference was robust to the sample quality variation estimated using RIN and PMI values, as well as to within-population variability estimated by subsampling individuals within populations. Furthermore, reanalysis of the lipidome data from the previous study based on 14 individuals (Khrameeva *et al.*, 2014) revealed HC-specific differences coinciding with the lipid abundance differences detected in our study.

The separation of adult HC individuals from WE and AA individuals with respect to lipid and polar metabolite abundance composition of the prefrontal cortex is a novel

observation, which contrasts with genetic and gene expression distances reported among populations. Specifically, the three populations used in the analysis are approximately equidistant from one another at the genome level, given the admixed genetic background of AA individuals (Parra, 2007; Tishkoff *et al.*, 2009). Similarly, no excess of HC- or east Asian-specific differences were reported by studies examining population-specific gene expression variation (Hughes *et al.*, 2015; Rawlings-Goss *et al.*, 2014). Nonetheless, some epigenetic effects, such as hypomethylation at *BRSK2*, were shown to be characteristic of East Asian individuals (Giuliani *et al.*, 2016). Furthermore, a concentration pattern specific to HC was shown at the lipidome level in a study analyzing whole blood composition in Chinese, Malay, and Indian individuals, although most lipid abundance differences were reported between individuals of Indian and non-Indian descent (Saw *et al.*, 2017).

The absence of pronounced lipid and metabolic differences distinguishing HC individuals during the first year of life, the period characterized by more uniform feeding and living routines, suggests that observed HC-specific differences might be environmental. On the other hand, studies examining dietary effects on different tissues, conducted in mice at the gene expression level (Bozek *et al.*, 2015) and in macaques at the lipid abundance level (Somel *et al.*, 2008), reported little or no detectable dietary effects in the brain, in contrast to non-neural tissues. Furthermore, the clustering of HC-specific lipid and metabolite concentration differences in particular functional pathways, detected in our study, might imply a possible link between these differences and variation in brain organization, functionality, and disease susceptibility among human populations

(Bai *et al.*, 2012; Hemming *et al.*, 2011; Tang *et al.*, 2018; Wright Willis *et al.*, 2010).

The exact connection between differences in lipid and polar metabolite abundance observed in our study and brain function or dysfunction needs further investigation, including evaluation of differences in lipid and polar metabolite composition between various cell types of the brain, as well as research on the link between functional properties of cellular membranes and the abundance of specific lipid compounds.

Due to the nature of the samples used in our study, we were unable to distinguish between the effect of environmental and genetic factors on the inter-population lipidome variation. However, decoupling the genetic and environmental effects for human populations, especially in studies involving postmortem tissue samples, represents a challenge. The WE cohort examined in our study did include samples from two locations, North America and Western Europe, but the environments at these locations are hardly distinct. Nonetheless, regardless of the cause, our study shows that the lipid and polar metabolite composition of the prefrontal cortex differs among populations, particularly in adult HC individuals.

The presence of population-specific features of the brain molecular organization has implications for further investigations, including a detailed analysis of the molecular brain composition across multiple human individuals. Additionally, our results provide a basis for the design of precision medicine studies, including clinical trial customization and treatment selection. Such studies are essential, given the multiple indications of population differences in brain morphology (Tang *et al.*, 2018), protein sequence

variation associated with lipid abundance (Lu *et al.*, 2017), and differential disease susceptibility (Xie *et al.*, 2017).

While the measurement of modern population metabolic differences is a straightforward task, the research on ancient population metabolome is challenging. Genomic changes in ancient humans could point the potential source of metabolic differences, however, as shown for modern populations, the metabolic alterations may not mirror genetic variations. Analysis of metabolic differences associated with evolution of human lineage might also shed the light, when the whole reconstruction of extinct populations metabolome is hardly achievable.

To find metabolic differences that set humans apart from their closest evolutionary relatives we investigated the metabolomes of the brain, muscle and kidney in humans, apes and monkeys. We find that steady state concentrations of many compounds involved in amino acid metabolism are present in higher or lower levels in humans than in other primates. In the future, it may be of interest to investigate the consequences of these human-specific metabolic features for the synthesis and catabolism of amino acids.

Oxidative phosphorylation and purine biosynthesis are presented in lower concentrations in humans than in the other primates analyzed. Whereas oxidative phosphorylation is lower in the three brain regions analyzed but not in muscle and kidneys, purine biosynthesis is decreased in all tissues analyzed, although most drastically in brain.

Humans and apes diverged so long ago that almost every gene carries changes that potentially alter its function by affecting its regulation or the structure of the encoded

gene product. In contrast, modern humans and Neanderthals and Denisovans diverged so recently that for about 90% of the genome, the two archaic human groups fall within the variation of present-day humans (Green *et al.*, 2010). Furthermore, when modern and archaic humans met about 50 000 years ago, they interbred, resulting in that in about half the genome, some present-day humans carry DNA from Neanderthals (Sankararaman *et al.*, 2014). The number of proteins that carry amino acid substitutions in all or almost all humans that differ from Neanderthals and apes is therefore only about one hundred (Pääbo, 2014). It is unclear if any of these substitutions have any functional consequences.

The alanine to valine substitution at position 429 in ADSL is one of this small number of substitutions. It affects a position that is conserved as alanine in most tetrapods suggesting that it may be of importance. Position 429 is also located only three positions away from position 426, where an arginine to histidine substitution causes the most common form of adenylosuccinase deficiency in present-day humans (Edery *et al.*, 2003; Kmoch, 2000; Maaswinkel-Mooij *et al.*, 1997; Marie *et al.*, 1999; Race, 2000). Further evidence suggesting that a change in *ADSL* may have been of importance in the evolution of modern humans comes from a screen for genomic regions that have experienced selective sweeps in humans after their split from Neanderthals but before the separation of Africans and Eurasians (Racimo, 2016). In this work, a genomic region centered around *ADSL* is among the top 20 candidate regions, although it contains also other genes. Furthermore, previous work has shown that the A429V substitution reduces the thermal stability of the ADSL protein *in vitro* (Van Laer *et al.*, 2018). The A429V

substitution is therefore an attractive candidate for having functional consequences in modern humans and we decided to analyze if it might be involved in the reduced purine biosynthesis seen in present-day humans by investigating the function of the ancestral, Neanderthal-like and the derived, modern human-like forms of ADSL *in vitro* and *in vivo*.

We confirm the previous finding (Van Laer *et al.*, 2018) that the A429V substitution does not affect the kinetic properties of the ADSL enzyme but decreases its thermal stability. We show that the substitution also decreases the stability of the tetrameric complex of the enzyme when exposed to a denaturing agent. When introduced in the mouse ADSL protein, this substitution similarly reduces the stability of the enzyme, either alone or in conjunction with a primate-specific substitution at the adjacent position 428. When the latter two substitutions are introduced into a mouse, it reduces the enzymatic activity detected in nine tissues analyzed, most drastically in the brain, and results in a reduction in purine biosynthesis, thus recapitulating differences seen between humans and chimpanzees and macaques.

To investigate how the A429V substitution may affect the metabolism of human cells, we used CRISPR-*Cas9* to introduce the ancestral, Neanderthal-like substitution into human cells. The concentrations of all nine metabolites detected downstream of ADSL in purine biosynthesis are increased in human cells carrying the ancestral substitution. In chimpanzee cells, the same nine metabolites occur in concentrations higher than in wild-type human cells. Notably, the expression of ADSL messenger RNA does not differ between human and chimpanzee cells, nor between wild type and ancestralized cells (not

shown). Thus, the A429V substitution is responsible for much or all of the difference in purine biosynthesis observed when human tissues are compared to ape and monkey tissues, indicating that this change in metabolism occurred in humans after their separation from the ancestor shared with Neanderthals and Denisovans.

An interesting question is what down-stream effects a reduced purine biosynthesis in modern humans might have. In this regard, it is intriguing that the A429V substitution results in a reduction in phosphorylation of AMPK, a major regulator of cellular energy homeostasis. Reduced AMPK activity may have numerous consequences. One of these may be to reduce oxidative phosphorylation, raising the possibility that the A429V substitution in ADSL may cause not only the reduction in purine biosynthesis seen in modern human tissues but also the reduction in oxidative phosphorylation through the effect of AMPK directly on this pathway (Lantier *et al.*, 2014; Nam *et al.*, 2016) and/or through its effects on the generation of mitochondria in cells (Bergeron *et al.*, 2001; Jager *et al.*, 2007; Marin *et al.*, 2017; Reznick *et al.*, 2007). However, we find no evidence that oxidative phosphorylation is affected in the ancestralized human cells nor in the mice carrying the human substitution. Thus, if the A429V substitution contributes also to a reduction in oxidative phosphorylation in humans, it must do so in conjunction with other genetic changes.

Indeed, although ADSL is expressed and functions in all tissues, it is interesting that the down-regulation of purine biosynthesis in humans relative to apes, and in humanized mice relative to wild-type mice, is most pronounced in the brain. It is also interesting that mutations in humans that affect enzymes involved in purine metabolism

have pathological consequences in the nervous system more than in other organs (Fumagalli *et al.*, 2017; Micheli *et al.*, 2011). It is thus possible that the A429V substitution in ADSL has contributed to human-specific changes in brain development and function. Future work will have to address this and other possibilities.

Along with other specific clinical disorders and neurological problems, disruption of purine biosynthesis was also associated with autistic features (Jinnah *et al.*, 2013). However, the exact cause of ASD remains unknown. Some researchers speculated that Neanderthal introgression into modern human genome could impact the ASD status (Mozzi *et al.*, 2017).

Our study also demonstrated the excess of metabolic differences clustering in purine metabolism in ASD. Moreover, the differences were identified to be human-specific; the observation aligns well with the hypothesis postulating disruption of recently evolved cognitive mechanisms in ASD (Liu *et al.*, 2016). Interestingly, ASD metabolite levels are intermediated between unaffected individuals and chimpanzees. The pattern mirrors the Neanderthal-like metabolite abundances being lower chimpanzees, but higher present-day humans abundances. Thus, ASD metabolome may inherit Neanderthal metabolome.

The presence of population-specific features of the molecular brain organization has implications for further investigations, including detailed analysis of the molecular brain composition across multiple human individuals. Additionally, our results provide a base for design of precision medicine studies, including clinical trials customization and treatment selection. Such studies are important, given multiple indications of population

differences in brain morphology (Tang *et al.*, 2018), protein sequence variation associated with lipid abundance (Lu *et al.*, 2017), as well as differential disease susceptibility, including the risk of hemorrhagic stroke (Xie *et al.*, 2017).

Chapter 5. Methods

Mass spectrometry methods

GC-MS measurements

Metabolites were extracted from the frozen tissue powder by a methanol:water:chloroform (2.5:1:1 (v/v/v)) extraction. In brief 100 mg of frozen powdered tissue material was resuspended in 1 mL extraction solution containing 0.1 $\mu\text{g mL}^{-1}$ of U- $^{13}\text{C}_6$ -sorbitol. The samples were incubated for 10 min at 4°C on an orbital shaker. This step was followed by ultrasonication in a bath-type sonicator for 10 min at room temperature. Finally the insoluble tissue material was pelleted by a centrifugation step (5 min; 14,000 g) and the supernatant transferred to a fresh 2 mL Eppendorf tube. To separate the organic from the aqueous phase 300 μL H₂O and 300 μL chloroform were added to the supernatant, vortexed and centrifuged (2 min; 14,000 g). Subsequently, 200 μL of the upper, aqueous phase were collected and concentrated to complete dryness in a speed vacuum at room temperature. Extract derivatization and GC-MS measurements were performed according to (Lisec *et al.*, 2006).

The obtained metabolite concentration values (apex height of the quantitative compound identifier mass) were normalized within each sample to the abundance of an internal standard (^{13}C sorbitol) and log₁₀ transformed. To avoid negative values, metabolite/standard ratios were scaled up by factor 3,000 prior to log₁₀ transformation.

CE-MS measurements

CE-MS measurements were conducted in PFC samples of 4 humans, 4 chimpanzees and 4 rhesus macaques. For each sample, metabolites were extracted from the frozen tissue powder by 1ml methanol containing 20 μ M each of L-Methionine sulfone, 2-Morpholinoethanesulfonic acid, monohydrate and sodium d-camphor-10-sulfonic acid. Then, 500 μ l of lysate was transferred to an Eppendorf tube containing 500 μ l chloroform and 200 μ l of Milli-Q water. 300 μ l of the aqueous phase, after 30 seconds of vortexing and 15 min of centrifugation at 4°C, was transferred to an ultrafiltration tube (Milipore). Concentrating the filtered liquid to complete dryness was then performed in a speed vacuum for 3 hours at 35°C. Just before the CE-MS analyses, the dried samples were mixed with 100 μ l of Milli-Q water containing 100 μ M each of 3-aminopyrrolidine and trimesate, and filtered with an ultrafiltration tube (Milipore) at 9,100 \times g for 2h at 4°C. 7 μ l of filtrate was used for the CE-MS analyses. CE-TOF-MS (Agilent Technologies) was then used to detect both cationic metabolites and anionic metabolites. The instrumentation and measurement conditions used for CE-TOF-MS were according to (Sugimoto *et al.*, 2012).

The in-house software MasterHands was used to perform peak detection, time alignment, and peak area integration. Concentrations of each metabolite in the samples were calculated based on the comparison of peak area normalized by internal standards' in the sample and external standard mixture. Metabolites with concentrations detected in more than half of the samples were included in the following analysis.

LC-MS measurements

In the LC-MS datasets, in addition to the individual cells samples, we measured mixtures of samples (pooled samples, EQC) after every 10th sample, providing us information on system performance (sensitivity and retention time consistency), sample reproducibility, and compound stability over the time of the MS-based analysis.

The dried extracts were resuspended in 100 μ l of ice-cold 20% aqueous solution of acetonitrile prior to mass spectrometry analysis. After brief rigorous vortexing the samples were incubated for 30 min at 4 °C on an orbital shaker followed by a 10 min ultra-sonication in an ice-cooled sonication bath and centrifugation for 10 min at 15.000 x g at 4 °C. For mass spectrometry analysis, 40 μ l of supernatant was transferred to a 350 μ l auto sampler glass vials (Glastechnik Grafenroda, Germany). Chromatography separation of metabolites prior to mass spectrometry was performed using Acquity I-Class UPLC system (Waters, UK). Metabolites were separated on a normal phase unbounded silica column RX-SIL (100 mm x 2.1 mm, 1.8 μ m, Agilent, US) coupled to a guard precolumn with the same phase parameters. The mobile phases used for the chromatographic separation were water containing 10 mM ammonium acetate, 0.2 mM ammonium hydroxide in water:acetonitrile (95: 5(v:v)) mixture (buffer A) (pH value 8.0) and 100% acetonitrile (buffer B). The gradient separation was: 0 min 0% buffer A, 0.01-15 min linear gradient from 0% to 100% buffer A, 15-18 min 100% buffer A, 18-19 min linear gradient from 100% buffer A to 0% buffer A, and 19-32 min 0% buffer A. After 1 min washing with 100% buffer A the column was re-equilibrated with 100% buffer B. The flow rate was set to 500 μ l/min. The column temperature was maintained at 32 °C. The mass spectra were acquired in positive and negative mode using a heated

electrospray ionization source in combination with Q Exactive Hybrid Quadrupole-Orbitrap mass spectrometer (Thermo Scientific, Germany). Negative ion mode samples were run after the positive ion mode cohort with 6 μ l injection of non-diluted samples. MS settings in positive acquisition mode: spray voltage was set to 4.5 kV in positive mode and to 3 kV in negative acquisition mode, S-lens RF level at 70, and heated capillary at 250 °C; aux gas heater temperature was set at 350 °C; sheath gas flow rate was set to 45 arbitrary units; aux gas flow rate was set to 10 arbitrary units; sweep gas flow rate was set to 4 arbitrary units. Full scan resolutions were set to 70 000 at m/z 200. Full scan target was 10e6 with a maximum fill time of 50 ms. The spectra were recorded using full scan mode, covering a mass range from 100– 1500 m/z.

For quality control (TQC), a pooled sample of all metabolic extracts was prepared and injected 4 times before initiating the runs in order to condition the column, at least 4 times after each sub-cohort, and after the completion of the runs. In addition, the TQC sample was injected every 48 sample injections to assess instrument stability and reproducibility.

Computational methods

Lipid compounds preprocessing

Data alignment and pre-processing of the lipid dataset were performed using the QI software (Version 2.2, www.nonlinear.com). Lipid peaks with liquid chromatography retention times shorter than 1.5 minutes or longer than 18 minutes were excluded from the analysis. The upper mass-to-charge ratio cutoff was set to 1400 Da. This cutoff includes the vast majority of lipid classes contained in tissues, with exception of low

abundance lipid classes containing four fatty acid residues, such as cardiolipins (Southam *et al.*, 2017). Only lipid peaks present with intensity greater than 200 in 75% of samples were kept in the analysis. For the remaining peaks, the intensity values smaller than 200 were replaced by the constant value equal to 200. Lipid peaks with the same mass-to-charge ratio observed at numerous different retention times were excluded from the analysis as potential contaminants.

We further excluded lipid peaks potentially confounded by their processing order during mass spectrometry measurements (run order). For each lipid peak, we fitted a support vector regression (svr) model with Gaussian kernel (sklearn.svm.SVR PYTHON module, parameters: C=100,000, epsilon 0.1, gamma 0.0001) to predict the peak intensity based on the run order, excluding three samples with the highest intensities for a more robust estimation. We then calculated the coefficient of determination (R^2) of this prediction. The top 300 lipid peaks detected in the positive ionization mode and 250 peaks detected the negative mode with strongest run order dependency were excluded from the downstream analysis ($R^2 > 0.41$ and > 0.47 for positive and negative modes respectively).

The intensities of lipid compounds retained after the above-mentioned filtration procedures were upper-quartile normalized and \log_2 transformed.

Polar compounds preprocessing

Polar metabolite data analysis was performed using the TargetSearch package according to (Cuadros-Inostroza *et al.*, 2009). Briefly, we describe the performed steps below. Settings for the peak-peaking and retention time alignment were as follows: m/z

range from 85 up to 750 units, intensity threshold 50 units, “smoothing” peak picking method, time window 10 sec. The compound annotation was performed using the Golm metabolome database (Hummel *et al.*, 2007), followed by the exclusion of un-annotated metabolites. Common contaminant masses 147-149 m/z were excluded, top 15 intensities from each library spectrum were selected, retention index windows were set to 2000, 1000, 200. Spearman correlation threshold = 0.95 was used for detection of correlating selective masses. The average retention time index of correlating selective masses was used for the calculation of compounds’ elution times. Multiple peaks corresponding to the same compound were collapsed based on the Spearman correlation threshold = 0.95 and elution time difference within 500 retention index units. For more details on the procedure, see (Cuadros-Inostroza *et al.*, 2009) and TargetSearch documentation.

The polar metabolites detected in less than 50% of the samples were excluded from the subsequent analyses. The remaining missing values were filled with the minimal intensity of the matrix. The \log_2 transformation of the polar metabolite intensities, linear regression for the experimental batch correction, and quantile normalization were applied subsequently to generate the table of polar metabolite intensities.

Data filtration

Both polar metabolites and lipids with intensities potentially affected by postmortem interval duration (PMI) were removed. The PMI effect was determined based on the Spearman correlation between the compound intensity values across samples and samples’ PMI (nominal p-value threshold <0.01). Because we expected strong effect of age factor on lipid and metabolite abundance, only samples of individuals with ages

greater than 5 years were used to estimate these correlations. The intensities of 1,670 lipids and 258 metabolites retained after above-mentioned procedure were used for all of the subsequent analysis, unless indicated otherwise. Additionally, part of the analysis was repeated using a more stringent exclusion criterion for potential PMI effects, where compounds were omitted from the analysis using a more relaxed p-value threshold of Spearman correlation between their intensities and PMI (nominal p-value threshold < 0.1).

To assess the effect of sample preservation quality on the results, an additional analysis of the population differences was conducted using a subset of 82 samples with high RNA preservation (RNA integrity number (RIN) ≥ 7).

Sample sets definitions

We defined datasets DS:0-4 and DS:5-71 as follows: samples with ages less than 5 years were assigned to DS:0-4 ($n = 74$), samples with ages greater than 5 years were assigned to DS:5-71 ($n = 229$). We defined six age groups A1-A6 as follows: A1 included samples from less than one-year-old individuals ($n = 24$), A2 included samples from 1-4 years-old individuals ($n = 50$), A3 included samples from 5-14 years-old individuals ($n = 41$), A4 included samples from 15-24 years-old individuals ($n = 56$), A5 included samples from 25-44 years-old individuals ($n = 56$), and A6 included samples from individuals with ages greater than full 44 years of age ($n = 76$) (Figure 4A).

T-distributed Stochastic Neighbor Embedding (t-SNE) analysis

The t-SNE analysis was conducted using “sklearn.manifold” PYTHON module

with the following parameters for lipid dataset: $n_components=2$, $perplexity=30$, $learning\ rate=100$, $metric='correlation'$, $early\ exaggeration=100$, $random_state=5$, and following parameters for metabolite dataset: $n_components=2$, $perplexity=30$, $learning\ rate=100$, $metric='correlation'$, $early\ exaggeration=12$, $random_state=0$. In addition to t-SNE analysis based on all detected lipids, the same analysis was also performed based on intensities of 900 annotated lipids.

Population specificity analysis

For both polar metabolites and lipids, to identify the significant intensity differences between three populations within DS:0-4 and DS:5-71 datasets, we subsampled equal number of individuals from each of the three populations within these datasets: 13 samples of each population in DS:0-4 and 25 samples of each population in DS:5-71. We then used t-test to compare the intensities in one population to the intensities in the other two populations combined. In each subsampling, compounds with t-test $p < 0.05$ after Benjamini-Hochberg correction were classified as population-specific. The subsampling procedure was performed 100 times to calculate the average number of population-specific differences for each of the three populations. To define HC-specific lipids and polar metabolites that were used in subsequent analysis, we performed the same procedure described above using the entire DS:5-71 dataset without subsampling.

We implemented a logistic regression model with lasso regularization to predict the population identity using DS:5-71 samples. Specifically, we randomly selected 31

samples from each population in the DS:5-71. The 93 samples selected from three populations were then randomly split into two parts. Two-thirds of the randomly selected samples ($n = 62$) were assigned as the training set, and the remaining one-third ($n = 31$) was assigned as the test set. Centering parameters (mean value of each compound) and scaling parameters (standard deviation of each compound) were estimated from the training set. Both training and test data were normalized according to these centering and scaling parameters. The logistic regression model was trained on the training set to separate one population from the other two combined using different hyperparameter C values (0.01, 0.1, 1, 10, 100, 500, 1000, 2500, 5000, 10000). Each time, the area under the receiver operating characteristic curve (ROC AUC) performance measure was calculated for the predictions of the test set. This procedure was repeated 100 times to estimate the average performance of the classifier on different test sets. Because the performance of the classifier did not depend strongly on the hyperparameter C (Figures 3C, 3D), we report performances for the arbitrarily chosen $C = 1000$ without the risk of overfitting the model to the data used for performance validation.

We defined HC-specific compounds, both polar metabolites and lipids, using stability selection procedure, as described in (Meinshausen *et al.*, 2008). Specifically, we randomly subsampled DS:5-71 individuals and split them into test and train sets, as described in the previous paragraph, followed by the construction of a HC-separating logistic regression model on the training set with lasso regularization and hyperparameter $C = 1000$. Next, we identified compounds selected by the model. We performed 10,000 iterations of this procedure to rank the compounds based on the number of iterations in

which they got selected by the predictive model. An arbitrary cutoff of 200 compounds for lipids and 50 compounds for polar metabolites was chosen to identify HC-specific compounds.

Population analysis within specified age groups

To analyze population divergence within specified age groups, we implemented classification-based and correlation-based approaches for both polar metabolite dataset and lipid dataset. Additionally, we selected age-unbiased population-distinguishing compounds that were not affected by the number of samples in the population and age groups. These compounds were used in the correlation-based analysis.

For the classification-based approach, we excluded the DS:0-4 samples (age groups A1 and A2) at each iteration and one sample from each of the A3-A6 age groups. Using the remaining samples, we performed stability selection and built a logistic regression using the 100 top compounds to predict the population identity of the samples excluded during the first step. We repeated this procedure until the population identity was predicted for each sample at least once and calculated the mean classification accuracy for each sample. We then used the median accuracy within a ten-sample-wide sliding window, with samples sorted according to age, to estimate the performance of the logistic regression model depending on the samples' age.

Age-unbiased population-distinguishing compounds were defined as follows. For each population pair and each age group, four samples per population were randomly chosen 1,000 times. For each subsampling iteration, we performed t-test for each population pair and age group, and selected the compounds showing positive or negative

differences with $p\text{-value} < 0.1$. The identified positive and negative differences were separately ranked based on the number of occurrences across the 1,000 subsampling iterations. We then selected the top-ranked 25 lipids and 10 polar metabolites showing positive differences for the corresponding population pair and age group, and same number of compounds showing negative differences. A union of these lipids and metabolites among all population comparisons and age groups was used to define the age-unbiased population-distinguishing lipid and metabolite sets.

To calculate the divergence of the three populations using correlation measurements, we randomly subsampled four samples in each A1-A6 age group 10,000 times and measured the Spearman correlation between the means of the selected samples. The analysis was conducted using age-unbiased population-distinguishing compounds and HC-specific compounds (defined using stability selection procedure).

Consistency analysis

To match the lipids from the current and the published datasets (Khrameeva *et al.*, 2014), the retention times were aligned using a select set of retention times anchor points and linear interpolation between them. Lipids were matched using 5 ppm threshold and 6 seconds retention time window. Only unique matches were retained. In this section, we refer to these lipids as “matched lipid compounds”.

To assess consistency of HC differences between current DS:5-71 and published dataset, for each dataset we calculated mean lipid intensity values of AA and WE populations samples combined and the mean lipid intensity values of HC population samples. Next, we calculated, for current and published datasets, the fold-changes of lipid

intensities between HC-population and the other two by calculating differences of above-mentioned mean values. We used Spearman correlation and HC-specific lipids contained in the set of matched lipid compounds to calculate correlation of these fold-changes between current and published datasets.

To assess the performance of the predictive model on an external dataset, we trained a predictive model using current dataset and predicted population identity of published dataset (Khrameeva *et al.*, 2014) samples, as follows. First, lipid intensities were normalized between experiments. To this purpose, 31 samples were selected from each population from DS:5-71. Mean and standard deviations were calculated for each lipid. Repeating this procedure 1,000 times, we calculated, for each lipid, an average centering (average of the mean values) and scaling (average of the standard deviation) value for the current dataset. Data were normalized according to these centering and scaling values. For published dataset, data were normalized using mean and standard deviation across samples. Because not all lipid predictors were present in published dataset and predictive power of a given lipid compound depends on the presence of other lipids in the model, we performed stability selection (Meinshausen *et al.*, 2008) as described in the Methods section, but restricting the procedure to matched lipid compounds. This produced a ranking for the matched lipids. Next, we built logistic regression model with $C = 1000$ and lasso normalization to train a predictive model on DS:5-71 samples and top-ranked lipids, and predicted population identity of published dataset samples. Using a varying amount of top-ranked lipids, we observed that the performance of the model peaked at 22 predictors.

Lipid annotation and enrichment analysis

Lipid annotation was performed using mass search with a tolerance of 5 ppm against the LIPID MAPS database (Sud *et al.*, 2007). The possible adducts were set to $[M+H]^+$, $[M+Na]^+$, $[M+NH_4]^+$ in positive ionization mode, and $[M-H]^-$, $[M-H+HCOOH]^-$, $[M-H+CH_3COOH]^-$ in negative ionization mode. For functional enrichment analysis, Kyoto Encyclopedia of Genes and Genomes (KEGG) database (Kanehisa *et al.*, 2017) was used to link lipids and polar metabolites to genes. We used hypergeometric test to assess, for each metabolic pathway, the enrichment of genes linked to HC-specific lipids and metabolites, compared to genes linked to all annotated lipids and metabolites.

Principal component analysis

We performed principal component analysis (PCA) and multidimensional scaling (MDS) for each mice and primates tissue separately in order to identify and to remove outlier samples.

Differential concentration test

We implemented t-test to determine metabolites with concentrations significantly different in mice and cell lines. A permutation procedure was used to assess the false positive rate. Briefly, we shuffled sample labels, applied t-test to two random groups of samples, and repeated this procedure 1000 times. Then, a permutation p-value was used to estimate the ratio of permutations resulting in equal or greater number of metabolites with significant concentration changes at a chosen nominal significance cutoff (0.05). To estimate the probability of increase and decrease of metabolite concentrations, we

performed binomial test. Human-specific metabolites were detected by both t-test and ANOVA analysis.

Total variation analysis

Percent of total variation explained by factors was estimated using analysis of variance (ANOVA) for lipid and metabolite datasets.

ANCOVA analysis

To identify metabolites with concentration differences between ASD samples and unaffected controls, we used an analysis of covariance (ANCOVA) as described in (Hyötyläinen *et al.*, 2014). Briefly, for each metabolite we chose the best polynomial regression model with age as a predictor and concentration values as a response based on adjusted R² criterion. Next, we used the F-test to evaluate whether the addition of disease/control status parameter significantly improved this model. The test was performed twice, using ASD samples as a reference for choosing the best polynomial regression model in one run, and control samples in the other. The resulting p-values were adjusted by the Benjamini-Hochberg (BH) approach. If the metabolite passed the BH corrected p-value threshold of 0.05 in both cases, the compound was classified as an ASD-related metabolite.

Chapter 6. Conclusions

Metabolomic and lipidomic studies complement genomics observations of human inter-population variability. However the postgenomic level signatures of variations are poorly explored. The results of this work reveal the presence of the metabolic and lipidomic differences among modern populations as well as differences between present-day and extinct populations in the brain. The study demonstrated that metabolic differences might not mirror genetic and epigenetic variation among individuals, but display additional differences. Although the existence of population differences in brain lipodome composition was suggested previously its sample size prevented detailed analysis of these differences. Here, we conducted a more extended survey of brain composition variation among human populations. In addition to covering more individuals, the study explored the dependence of the population-specific differences on individuals' age and included polar metabolites, thus assessing a level of molecular brain organization not explored in previous studies.

Furthermore, in my thesis I focus on metabolic differences that influence the biology of modern humans. By analyzing the metabolomes of muscle, kidney and three different regions of the brain from humans, chimpanzees and macaques, we find that many aspects of amino acid metabolism differ between humans and the other two primates in all tissues analyzed. Among metabolic pathways, oxidative phosphorylation is less active in the brains of humans than the other primates and purine biosynthesis is less active in the human brain as well as in other tissues. In purine biosynthesis, we find that metabolites downstream of the enzyme adenylosuccinate lyase occur at lower concentrations in

humans than in the other primates. ADSL carries an amino acid substitution that is unique to modern humans relative to apes and Neanderthals and Denisovans and has been shown to affect the stability of the enzyme. By introducing the modern human-like substitution in the genome of mice, and the ancestral, Neanderthal-like substitution in the genomes of human cells, we show that this substitution is responsible for much or all of this metabolic change in present-day humans. We show the potential power of model systems, such as transgenic mice and modified human cell lines to study the metabolic effects of the modern human-specific alterations. Being the first metabolic reconstruction of modern humans molecular phenotype features, the results provide ideas for the future work of the human brain development and function.

Bibliography

- Abarca, Jorge, & Bustos, Gonzalo. (1999). Differential regulation of glutamate, aspartate and γ -amino-butyrate release by N-methyl-d-aspartate receptors in rat striatum after partial and extensive lesions to the nigro-striatal dopamine pathway. *Neurochemistry International*, 35(1), 19–33. [https://doi.org/10.1016/S0197-0186\(99\)00029-7](https://doi.org/10.1016/S0197-0186(99)00029-7)
- Abbott, Sarah K., Jenner, Andrew M., Spiro, Adena S., Batterham, Marijka, Halliday, Glenda M., & Garner, Brett. (2015). Fatty acid composition of the anterior cingulate cortex indicates a high susceptibility to lipid peroxidation in Parkinson's disease. *Journal of Parkinson's Disease*. <https://doi.org/10.3233/JPD-140479>
- Adeva-Andany, M., López-Ojén, M., Funcasta-Calderón, R., Ameneiros-Rodríguez, E., Donapetry-García, C., Vila-Altesor, M., & Rodríguez-Seijas, J. (2014). Comprehensive review on lactate metabolism in human health. *Mitochondrion*. <https://doi.org/10.1016/j.mito.2014.05.007>
- Bai, Jordan, Abdul-Rahman, Muhammad Farid, Rifkin-Graboi, Anne, Chong, Yap Seng, Kwek, Kenneth, Saw, Seang Mei, ... Qiu, Anqi. (2012). Population Differences in Brain Morphology and Microstructure among Chinese, Malay, and Indian Neonates. *PLoS ONE*, 7(10). <https://doi.org/10.1371/journal.pone.0047816>
- Bergeron, Raynald, Ren, Jian Ming, Cadman, Kevin S., Moore, Irene K., Perret, Pascale, Pypaert, Marc, ... Shulman, Gerald I. (2001). Chronic activation of AMP kinase results in NRF-1 activation and mitochondrial biogenesis. *American Journal of Physiology-Endocrinology and Metabolism*, 281(6), E1340–E1346. <https://doi.org/10.1152/ajpendo.2001.281.6.E1340>

- Bogner, W., Gruber, S., Trattnig, S., & Chmelik, M. (2012). High-resolution mapping of human brain metabolites by free induction decay ^1H MRSI at 7T. *NMR in Biomedicine*. <https://doi.org/10.1002/nbm.1805>
- Botchkarev, Vladimir A., & Fessing, Michael Y. (2005). Edar signaling in the control of hair follicle development. *The Journal of Investigative Dermatology. Symposium Proceedings / the Society for Investigative Dermatology, Inc. [and] European Society for Dermatological Research*, 10(3), 247–251. <https://doi.org/10.1111/j.1087-0024.2005.10129.x>
- Bozek, Katarzyna, Wei, Yuning, Yan, Zheng, Liu, Xiling, Xiong, Jieyi, Sugimoto, Masahiro, ... Khaitovich, Philipp. (2015). Organization and Evolution of Brain Lipidome Revealed by Large-Scale Analysis of Human, Chimpanzee, Macaque, and Mouse Tissues. *Neuron*, 85(4), 695–702. <https://doi.org/10.1016/j.neuron.2015.01.003>
- Brady, S. T., Siegel, G. J., Albers, R. W., & Price, D. L. (2012). *Basic Neurochemistry*. <https://doi.org/10.1016/C2009-0-00066-X>
- Chang, Shie Hong, Jobling, Stephanie, Brennan, Keith, & Headon, Denis J. (2009). Enhanced Edar Signalling Has Pleiotropic Effects on Craniofacial and Cutaneous Glands. *PLoS ONE*, 4(10), e7591. <https://doi.org/10.1371/journal.pone.0007591>
- Chang, Xiangyun, Li, Siyuan, Li, Jun, Yin, Liang, Zhou, Ting, Zhang, Chen, ... Sun, Kan. (2014). Ethnic Differences in MicroRNA-375 Expression Level and DNA Methylation Status in Type 2 Diabetes of Han and Kazak Populations. *Journal of Diabetes Research*, 2014, 1–7. <https://doi.org/10.1155/2014/761938>

- Chen, Chi-Ming A., Stanford, Arielle D., Mao, Xiangling, Abi-Dargham, Anissa, Shungu, Dikoma C., Lisanby, Sarah H., ... Kegeles, Lawrence S. (2014). GABA level, gamma oscillation, and working memory performance in schizophrenia. *NeuroImage: Clinical*, 4, 531–539. <https://doi.org/10.1016/j.nicl.2014.03.007>
- Chin, Alexander L., Negash, Selamawit, & Hamilton, Roy. (2011). Diversity and Disparity in Dementia. *Alzheimer Disease & Associated Disorders*, 25(3), 187–195. <https://doi.org/10.1097/WAD.0b013e318211c6c9>
- Crabtree, Gregg W., & Gogos, Joseph A. (2018). Role of Endogenous Metabolite Alterations in Neuropsychiatric Disease. *ACS Chemical Neuroscience*, 9(9), 2101–2113. <https://doi.org/10.1021/acchemneuro.8b00145>
- Crabtree, Gregg W., Park, Alan J., Gordon, Joshua A., & Gogos, Joseph A. (2016). Cytosolic Accumulation of L-Proline Disrupts GABA-Ergic Transmission through GAD Blockade. *Cell Reports*, 17(2), 570–582. <https://doi.org/10.1016/j.celrep.2016.09.029>
- Cuadros-Inostroza, Álvaro, Caldana, Camila, Redestig, Henning, Kusano, Miyako, Lise, Jan, Peña-Cortés, Hugo, ... Hannah, Matthew A. (2009). TargetSearch - a Bioconductor package for the efficient preprocessing of GC-MS metabolite profiling data. *BMC Bioinformatics*, 10(1), 428. <https://doi.org/10.1186/1471-2105-10-428>
- Dasgupta, Amitava, & Wahed, Amer. (2014). Drugs of Abuse Testing. In *Clinical Chemistry, Immunology and Laboratory Quality Control* (pp. 289–306). <https://doi.org/10.1016/B978-0-12-407821-5.00016-4>
- Davletov, Bazbek, & Montecucco, Cesare. (2010). Lipid function at synapses. *Current*

Opinion in Neurobiology, 20(5), 543–549.

<https://doi.org/10.1016/j.conb.2010.06.008>

Dennis, E., Rhee, S., Billah, M., & Hannun, Y. (1991). Role of phospholipases in generating lipid second messengers in signal transduction. *FASEB Journal*.

<https://doi.org/10.1096/fasebj.5.7.1901288>

Dwyer, D. S., Bradley, R. J., Kablinger, A. S., & Freeman, A. M. (2001). Glucose metabolism in relation to schizophrenia and antipsychotic drug treatment. *Annals of Clinical Psychiatry : Official Journal of the American Academy of Clinical Psychiatrists*, 13(2), 103–113. Retrieved from

<http://www.ncbi.nlm.nih.gov/pubmed/11534925>

Early, Dawnté R., Widaman, Keith F., Harvey, Danielle, Beckett, Laurel, Park, Lovingly Quitania, Farias, Sarah Tomaszewski, ... Mungas, Dan. (2013). Demographic predictors of cognitive change in ethnically diverse older persons. *Psychology and Aging*, 28(3), 633–645. <https://doi.org/10.1037/a0031645>

Ederly, P., Chabrier, S., Ceballos-Picot, I., Marie, S., Vincent, M., & Tardieu, M. (2003).

Intrafamilial variability in the phenotypic expression of adenylosuccinate lyase deficiency: A report on three patients. *American Journal of Medical Genetics*, 120A(2), 185–190. <https://doi.org/10.1002/ajmg.a.20176>

Ekroos, Kim, Jänis, Minna, Tarasov, Kirill, Hurme, Reini, & Laaksonen, Reijo. (2010).

Lipidomics: A tool for studies of atherosclerosis. *Current Atherosclerosis Reports*. <https://doi.org/10.1007/s11883-010-0110-y>

Erbslöh, F., Bernsmeier, A., & Hillesheim, H. R. (1958). Der Glucoseverbrauch des

Gehirns und seine Abhängigkeit von der Leber. *Archiv Für Psychiatrie Und Nervenkrankheiten Vereinigt Mit Zeitschrift Für Die Gesamte Neurologie Und Psychiatrie*. <https://doi.org/10.1007/BF00344388>

Fahy, Eoin, Subramaniam, Shankar, Brown, H. Alex, Glass, Christopher K., Merrill, Alfred H., Murphy, Robert C., ... Dennis, Edward A. (2005). A comprehensive classification system for lipids. *Journal of Lipid Research*. <https://doi.org/10.1194/jlr.E400004-JLR200>

Fietz, Simone A., & Huttner, Wieland B. (2011). Cortical progenitor expansion, self-renewal and neurogenesis—a polarized perspective. *Current Opinion in Neurobiology*, *21*(1), 23–35. <https://doi.org/10.1016/j.conb.2010.10.002>

Fon, E. A., Demczuk, S., Delattre, O., Thomas, G., & Rouleau, G. A. (1993). Mapping of the human adenylosuccinate lyase (*Adsl*) gene to chromosome 22q13.1→q13.2. *Cytogenetic and Genome Research*. <https://doi.org/10.1159/000133575>

Fraser, Hunter B., Lam, Lucia L., Neumann, Sarah M., & Kobor, Michael S. (2012). Population-specificity of human DNA methylation. *Genome Biology*, *13*(2). <https://doi.org/10.1186/gb-2012-13-2-r8>

Fu, X., Giavalisco, P., Liu, X., Catchpole, G., Fu, N., Ning, Z. B., ... Khaitovich, P. (2011). Rapid metabolic evolution in human prefrontal cortex. *Proceedings of the National Academy of Sciences*, *108*(15), 6181–6186. <https://doi.org/10.1073/pnas.1019164108>

Fumagalli, Marta, Lecca, Davide, Abbracchio, Maria P., & Ceruti, Stefania. (2017). Pathophysiological Role of Purines and Pyrimidines in Neurodevelopment:

Unveiling New Pharmacological Approaches to Congenital Brain Diseases.

Frontiers in Pharmacology, 8. <https://doi.org/10.3389/fphar.2017.00941>

Gallego Romero, Irene, Pai, Athma A., Tung, Jenny, & Gilad, Yoav. (2014). RNA-seq: Impact of RNA degradation on transcript quantification. *BMC Biology*.

<https://doi.org/10.1186/1741-7007-12-42>

Giuliani, Cristina, Sazzini, Marco, Bacalini, Maria Giulia, Pirazzini, Chiara, Marasco, Elena, Fontanesi, Elisa, ... Garagnani, Paolo. (2016). Epigenetic Variability across Human Populations: A Focus on DNA Methylation Profiles of the KRTCAP3 , MAD1L1 and BRSK2 Genes. *Genome Biology and Evolution*, 8(9), 2760–2773.

<https://doi.org/10.1093/gbe/evw186>

Gohlke, R. S. (1959). Time-of-Flight Mass Spectrometry and Gas-Liquid Partition Chromatography. *Analytical Chemistry*, 31(4), 535–541.

<https://doi.org/10.1021/ac50164a024>

Goldberg, D., Johnston, J., Cameron, S., Fletcher, C., Stewart, M., McMenamin, J., ... Raeside, F. (2001). Glucose metabolism in relation to schizophrenia and antipsychotic drug treatment. *Annals of Clinical Psychiatry*, 13(2), 103–113.

<https://doi.org/10.1023/A:1016671725396>

Görke, Robert, Meyer-Bäse, Anke, Wagner, Dorothea, He, Huan, Emmett, Mark R., & Conrad, Charles A. (2010). Determining and interpreting correlations in lipidomic networks found in glioblastoma cells. *BMC Systems Biology*.

<https://doi.org/10.1186/1752-0509-4-126>

Graessler, Juergen, Schwudke, Dominik, Schwarz, Peter E. H., Herzog, Ronny,

- Schevchenko, Andrej, & Bornstein, Stefan R. (2009). Top-down lipidomics reveals ether lipid deficiency in blood plasma of hypertensive patients. *PLoS ONE*.
<https://doi.org/10.1371/journal.pone.0006261>
- Grant, Gregor H., & Butt, Wilfrid R. (1970). Immunochemical methods in clinical chemistry. *Advances in Clinical Chemistry*. [https://doi.org/10.1016/S0065-2423\(08\)60390-X](https://doi.org/10.1016/S0065-2423(08)60390-X)
- Green, R. E., Krause, J., Briggs, A. W., Maricic, T., Stenzel, U., Kircher, M., ... Paabo, S. (2010). A Draft Sequence of the Neandertal Genome. *Science*, 328(5979), 710–722. <https://doi.org/10.1126/science.1188021>
- Green, & Tzagoloff, A. (1966). Role of lipids in the structure and function of biological membranes. *Journal of Lipid Research*.
- Gundogdu, Betul, Al-Lahham, Tawfiq, Kadlubar, Fred, Spencer, Horace, & Rudnicki, Stacy A. (2014). Racial differences in motor neuron disease. *Amyotrophic Lateral Sclerosis and Frontotemporal Degeneration*, 15(1–2), 114–118.
<https://doi.org/10.3109/21678421.2013.837930>
- Han, Tingting, Ren, Xingxing, Jiang, Dongdong, Zheng, Shuang, Chen, Yawen, Qiu, Huiying, ... Hu, Yaomin. (2018). Pathophysiological changes after lipopolysaccharide-induced acute inflammation in a type 2 diabetic rat model versus normal controls. *Diabetes Research and Clinical Practice*.
<https://doi.org/10.1016/j.diabres.2018.02.012>
- Hemming, J. Patrick, Gruber-Baldini, Ann L., Anderson, Karen E., Fishman, Paul S., Reich, Stephen G., Weiner, William J., & Shulman, Lisa M. (2011). Racial and

socioeconomic disparities in parkinsonism. *Archives of Neurology*, 68(4), 498–503.
<https://doi.org/10.1001/archneurol.2010.326>

Hughes, David A., Kircher, Martin, He, Zhisong, Guo, Song, Fairbrother, Genevieve L., Moreno, Carlos S., ... Stoneking, Mark. (2015). Evaluating intra- and inter-individual variation in the human placental transcriptome. *Genome Biology*, 16(1), 1–18. <https://doi.org/10.1186/s13059-015-0627-z>

Hummel, Jan, Selbig, Joachim, Walther, Dirk, & Kopka, Joachim. (2007). *The Golm Metabolome Database: a database for GC-MS based metabolite profiling*.
https://doi.org/10.1007/4735_2007_0229

Hunter, Mandana, Demarais, Nicholas J., Faull, Richard L. M., Grey, Angus C., & Curtis, Maurice A. (2018). Layer-specific lipid signatures in the human subventricular zone demonstrated by imaging mass spectrometry. *Scientific Reports*.
<https://doi.org/10.1038/s41598-018-20793-4>

Hyötyläinen, Tuulia, & Orešič, Matej. (2014). Systems biology strategies to study lipidomes in health and disease. *Progress in Lipid Research*, 55, 43–60.
<https://doi.org/10.1016/j.plipres.2014.06.001>

Jager, S., Handschin, C., St.-Pierre, J., & Spiegelman, B. M. (2007). AMP-activated protein kinase (AMPK) action in skeletal muscle via direct phosphorylation of PGC-1. *Proceedings of the National Academy of Sciences*, 104(29), 12017–12022.
<https://doi.org/10.1073/pnas.0705070104>

James, K. D. (2016). Animal Metabolites: From Amphibians, Reptiles, Aves/Birds, and Invertebrates. In *Pharmacognosy: Fundamentals, Applications and Strategy*.

<https://doi.org/10.1016/B978-0-12-802104-0.00019-6>

Jinnah, H. A., Sabina, Richard I., & Van Den Berghe, Georges. (2013). *Metabolic disorders of purine metabolism affecting the nervous system*.

<https://doi.org/10.1016/B978-0-444-59565-2.00052-6>

Jurecka, Agnieszka, Jurkiewicz, Elzbieta, & Tylki-Szymanska, Anna. (2012). Magnetic resonance imaging of the brain in adenylosuccinate lyase deficiency: a report of seven cases and a review of the literature. *European Journal of Pediatrics*, 171(1), 131–138. <https://doi.org/10.1007/s00431-011-1503-9>

Jurecka, Agnieszka, Zikanova, Marie, Kmoch, Stanislav, & Tylki-Szymańska, Anna. (2015). Adenylosuccinate lyase deficiency. *Journal of Inherited Metabolic Disease*, 38(2), 231–242. <https://doi.org/10.1007/s10545-014-9755-y>

Kanehisa, Minoru, Furumichi, Miho, Tanabe, Mao, Sato, Yoko, & Morishima, Kanae. (2017). KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Research*, 45(D1), D353–D361. <https://doi.org/10.1093/nar/gkw1092>

Khrameeva, Ekaterina E., Bozek, Katarzyna, He, Liu, Yan, Zheng, Jiang, Xi, Wei, Yuning, ... Khaitovich, Philipp. (2014). Neanderthal ancestry drives evolution of lipid catabolism in contemporary Europeans. *Nature Communications*, 5(1), 3584. <https://doi.org/10.1038/ncomms4584>

Kmoch, S. (2000). Human adenylosuccinate lyase (ADSL), cloning and characterization of full-length cDNA and its isoform, gene structure and molecular basis for ADSL deficiency in six patients. *Human Molecular Genetics*. <https://doi.org/10.1093/hmg/9.10.1501>

- Kokovay, Erzsebet, Wang, Yue, Kusek, Gretchen, Wurster, Rachel, Lederman, Patty, Lowry, Natalia, ... Temple, Sally. (2012). VCAM1 Is Essential to Maintain the Structure of the SVZ Niche and Acts as an Environmental Sensor to Regulate SVZ Lineage Progression. *Cell Stem Cell*, 11(2), 220–230.
<https://doi.org/10.1016/j.stem.2012.06.016>
- Kurochkin, Ilia, Khrameeva, Ekaterina, Tkachev, Anna, Stepanova, Vita, Vanyushkina, Anna, Stekolshchikova, Elena, ... Khaitovich, Philipp. (2019). Metabolome signature of autism in the human prefrontal cortex. *Communications Biology*.
<https://doi.org/10.1038/s42003-019-0485-4>
- Lantier, Louise, Fentz, Joachim, Mounier, Rémi, Leclerc, Jocelyne, Treebak, Jonas T., Pehmøller, Christian, ... Viollet, Benoit. (2014). AMPK controls exercise endurance, mitochondrial oxidative capacity, and skeletal muscle integrity. *The FASEB Journal*, 28(7), 3211–3224. <https://doi.org/10.1096/fj.14-250449>
- Lappalainen, Tuuli, Sammeth, Michael, Friedländer, Marc R., 't Hoen, Peter A. C., Monlong, Jean, Rivas, Manuel A., ... Dermitzakis, Emmanouil T. (2013). Transcriptome and genome sequencing uncovers functional variation in humans. *Nature*, 501(7468), 506–511. <https://doi.org/10.1038/nature12531>
- Lee, Peychii, & Colman, Roberta F. (2007). Expression, purification, and characterization of stable, recombinant human adenylosuccinate lyase. *Protein Expression and Purification*, 51(2), 227–234. <https://doi.org/10.1016/j.pep.2006.07.023>
- Li, Qian, Bozek, Katarzyna, Xu, Chuan, Guo, Yanan, Sun, Jing, Pääbo, Svante, ... Khaitovich, Philipp. (2017). Changes in Lipidome Composition during Brain

Development in Humans, Chimpanzees, and Macaque Monkeys. *Molecular Biology and Evolution*, 34(5), 1155–1166. <https://doi.org/10.1093/molbev/msx065>

Lisec, Jan, Schauer, Nicolas, Kopka, Joachim, Willmitzer, Lothar, & Fernie, Alisdair R. (2006). Gas chromatography mass spectrometry–based metabolite profiling in plants. *Nature Protocols*, 1(1), 387–396. <https://doi.org/10.1038/nprot.2006.59>

Liu, Xiling, Han, Dingding, Somel, Mehmet, Jiang, Xi, Hu, Haiyang, Guijarro, Patricia, ... Khaitovich, Philipp. (2016). Disruption of an Evolutionarily Novel Synaptic Expression Pattern in Autism. *PLOS Biology*, 14(9), e1002558. <https://doi.org/10.1371/journal.pbio.1002558>

Livingstone, F. B. (1984). The Duffy blood groups, vivax malaria, and malaria selection in human populations: a review. *Human Biology*, 56(3), 413–425. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/6386656>

Lohmueller, Kirk E., Indap, Amit R., Schmidt, Steffen, Boyko, Adam R., Hernandez, Ryan D., Hubisz, Melissa J., ... Bustamante, Carlos D. (2008). Proportionally more deleterious genetic variation in European than in African populations. *Nature*, 451(7181), 994–997. <https://doi.org/10.1038/nature06611>

Lombard, Julian H. (2006). A novel mechanism for regulation of retinal blood flow by lactate: Gap junctions, hypoxia, and pericytes. *American Journal of Physiology - Heart and Circulatory Physiology*. <https://doi.org/10.1152/ajpheart.01268.2005>

Lu, Xiangfeng, Peloso, Gina M., Liu, Dajiang J., Wu, Ying, Zhang, He, Zhou, Wei, ... Willer, Cristen J. (2017). Exome chip meta-analysis identifies novel loci and East Asian–specific coding variants that contribute to lipid levels and coronary artery

- disease. *Nature Genetics*, 49(12), 1722–1730. <https://doi.org/10.1038/ng.3978>
- Lukiw, W. J. (2005). A role for docosahexaenoic acid-derived neuroprotectin D1 in neural cell survival and Alzheimer disease. *Journal of Clinical Investigation*, 115(10), 2774–2783. <https://doi.org/10.1172/JCI25420>
- Maaswinkel-Mooij, P. D., Laan, L. A. E. M., Onkenhout, W., Brouwer, O. F., Jaeken, J., & Poorthuis, B. J. H. M. (1997). Adenylosuccinase deficiency presenting with epilepsy in early infancy. *Journal of Inherited Metabolic Disease*. <https://doi.org/10.1023/A:1005323512982>
- Makgoba, M., Savvidou, MD, & Steer, PJ. (2012). An analysis of the interrelationship between maternal age, body mass index and racial origin in the development of gestational diabetes mellitus. *BJOG: An International Journal of Obstetrics & Gynaecology*, 119(3), 276–282. <https://doi.org/10.1111/j.1471-0528.2011.03156.x>
- Marie, S., Race, V., Nassogne, M. C., Vincent, M. F., & Van den Berghe, G. (2002). Mutation of a Nuclear Respiratory Factor 2 Binding Site in the 5' Untranslated Region of the ADSL Gene in Three Patients with Adenylosuccinate Lyase Deficiency. *The American Journal of Human Genetics*, 71(1), 14–21. <https://doi.org/10.1086/341036>
- Marie, Sandrine, Cuppens, Harry, Heuterspreute, Michel, Jaspers, Martine, Tola, Eduardo Zambrano, Gu, Xiao Xiao, ... Van Den Berghe, Georges. (1999). Mutation analysis in adenylosuccinate lyase deficiency: Eight novel mutations in the re-evaluated full ADSL coding sequence. *Human Mutation*. [https://doi.org/10.1002/\(SICI\)1098-1004\(1999\)13:3<197::AID-HUMU3>3.0.CO;2-D](https://doi.org/10.1002/(SICI)1098-1004(1999)13:3<197::AID-HUMU3>3.0.CO;2-D)

- Marie, Sandrine, Heron, Bénédicte, Bitoun, Pierre, Timmerman, Thérèse, Van den Berghe, Georges, & Vincent, Marie-Françoise. (2004). AICA-Ribosiduria: A Novel, Neurologically Devastating Inborn Error of Purine Biosynthesis Caused by Mutation of ATIC. *The American Journal of Human Genetics*, *74*(6), 1276–1281.
<https://doi.org/10.1086/421475>
- Marin, Traci L., Gongol, Brendan, Zhang, Fan, Martin, Marcy, Johnson, David A., Xiao, Han, ... Shyy, John Y. J. (2017). AMPK promotes mitochondrial biogenesis and function by phosphorylating the epigenetic factors DNMT1, RBBP7, and HAT1. *Science Signaling*, *10*(464), eaaf7478. <https://doi.org/10.1126/scisignal.aaf7478>
- Masel, Meredith C., & Peek, M. Kristen. (2009). Ethnic Differences in Cognitive Function Over Time. *Annals of Epidemiology*, *19*(11), 778–783.
<https://doi.org/10.1016/j.annepidem.2009.06.008>
- Meinshausen, Nicolai, & Buehlmann, Peter. (2008). *Stability Selection*. Retrieved from <http://arxiv.org/abs/0809.2932>
- Mergenthaler, Philipp, Lindauer, Ute, Dienel, Gerald A., & Meisel, Andreas. (2013). Sugar for the brain: The role of glucose in physiological and pathological brain function. *Trends in Neurosciences*. <https://doi.org/10.1016/j.tins.2013.07.001>
- Micheli, Vanna, Camici, Marcella, G. Tozzi, Maria, L. Ipata, Piero, Sestini, Sylvia, Bertelli, Matteo, & Pompucci, Giuseppe. (2011). Neurological Disorders of Purine and Pyrimidine Metabolism. *Current Topics in Medicinal Chemistry*, *11*(8), 923–947. <https://doi.org/10.2174/156802611795347645>
- Min, Hye Kyeong, Lim, Sangsoo, Chung, Bong Chul, & Moon, Myeong Hee. (2011).

Shotgun lipidomics for candidate biomarkers of urinary phospholipids in prostate cancer. *Analytical and Bioanalytical Chemistry*. <https://doi.org/10.1007/s00216-010-4290-7>

Mozzi, Alessandra, Forni, Diego, Cagliani, Rachele, Pozzoli, Uberto, Clerici, Mario, & Sironi, Manuela. (2017). Distinct selective forces and Neanderthal introgression shaped genetic diversity at genes involved in neurodevelopmental disorders. *Scientific Reports*, 7(1), 6116. <https://doi.org/10.1038/s41598-017-06440-4>

Nam, Kee Soo, Sunhwa, O. H., & Shin, Incheol. (2016). Ablation of CD44 induces glycolysis-To-oxidative phosphorylation transition via modulation of the c-Src-Akt-LKB1-AMPK α pathway. *Biochemical Journal*. <https://doi.org/10.1042/BCJ20160613>

Nassogne, Marie Cécile, Henrot, Brigitte, Aubert, Geneviève, Bonnier, Christine, Marie, Sandrine, Saint-Martin, Christine, ... Vincent, Marie Françoise. (2000). Adenylosuccinase deficiency: An unusual cause of early-onset epilepsy associated with acquired microcephaly. *Brain and Development*. [https://doi.org/10.1016/S0387-7604\(00\)00154-6](https://doi.org/10.1016/S0387-7604(00)00154-6)

Nielsen, Rasmus, Akey, Joshua M., Jakobsson, Mattias, Pritchard, Jonathan K., Tishkoff, Sarah, & Willerslev, Eske. (2017). Tracing the peopling of the world through genomics. *Nature*, 541(7637), 302–310. <https://doi.org/10.1038/nature21347>

Oxford Dictionary of Biochemistry and Molecular Biology. (2007). *Choice Reviews Online*. <https://doi.org/10.5860/choice.44-3627>

Pääbo, Svante. (2014). The Human Condition—A Molecular Approach. *Cell*, 157(1),

216–226. <https://doi.org/10.1016/j.cell.2013.12.036>

Parra, Esteban J. (2007). Admixture in North America. *Pharmacogenomics in Admixed Populations*, 1–19. https://doi.org/10.1300/J043v12n02_05

Pietiläinen, Kirsi H., Sysi-Aho, Marko, Rissanen, Aila, Seppänen-Laakso, Tuulikki, Yki-Järvinen, Hannele, Kaprio, Jaakko, & Orešič, Matej. (2007). Acquired obesity is associated with changes in the serum lipidomic profile independent of genetic effects - A monozygotic twin study. *PLoS ONE*.
<https://doi.org/10.1371/journal.pone.0000218>

Prüfer, Kay, Racimo, Fernando, Patterson, Nick, Jay, Flora, Sankararaman, Sriram, Sawyer, Susanna, ... Pääbo, Svante. (2014). The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature*, 505(7481), 43–49.
<https://doi.org/10.1038/nature12886>

Psychogios, Nikolaos, Hau, David D., Peng, Jun, Guo, An Chi, Mandal, Rupasri, Bouatra, Souhaila, ... Wishart, David S. (2011). The human serum metabolome. *PLoS ONE*, 6(2). <https://doi.org/10.1371/journal.pone.0016957>

Race, V. (2000). Clinical, biochemical and molecular genetic correlations in adenylosuccinate lyase deficiency. *Human Molecular Genetics*, 9(14), 2159–2165.
<https://doi.org/10.1093/hmg/9.14.2159>

Racimo, Fernando. (2016). Testing for Ancient Selection Using Cross-population Allele Frequency Differentiation. *Genetics*, 202(2), 733–750.
<https://doi.org/10.1534/genetics.115.178095>

Rawlings-Goss, Renata A., Campbell, Michael C., & Tishkoff, Sarah A. (2014). Global

population-specific variation in miRNA associated with cancer risk and clinical biomarkers. *BMC Medical Genomics*, 7(1), 53. <https://doi.org/10.1186/1755-8794-7-53>

Ray, Stephen P., Deaton, Michelle K., Capodagli, Glenn C., Calkins, Lauren A. F., Sawle, Lucas, Ghosh, Kingshuk, ... Pegan, Scott D. (2012). Structural and Biochemical Characterization of Human Adenylosuccinate Lyase (ADSL) and the R303C ADSL Deficiency-Associated Mutation. *Biochemistry*, 51(33), 6701–6713. <https://doi.org/10.1021/bi300796y>

Reznick, Richard M., Zong, Haihong, Li, Ji, Morino, Katsutaro, Moore, Irene K., Yu, Hannah J., ... Shulman, Gerald I. (2007). Aging-Associated Reductions in AMP-Activated Protein Kinase Activity and Mitochondrial Biogenesis. *Cell Metabolism*, 5(2), 151–156. <https://doi.org/10.1016/j.cmet.2007.01.008>

Romualdi, Chiara, Balding, David, Nasidze, Ivane S., Risch, Gregory, Robichaux, Myles, Sherry, Stephen T., ... Barbujani, Guido. (2002). Patterns of human diversity, within and among continents, inferred from biallelic DNA polymorphisms. *Genome Research*, 12(4), 602–612. <https://doi.org/10.1101/gr.214902>

Rosenberg, N. A. (2002). Genetic Structure of Human Populations. *Science*, 298(5602), 2381–2385. <https://doi.org/10.1126/science.1078311>

Rouser, George, & Yamamoto, Akira. (1968). Curvilinear regression course of human brain lipid composition changes with age. *Lipids*, 3(3), 284–287. <https://doi.org/10.1007/BF02531202>

Sabeti, Pardis C., Varilly, Patrick, Fry, Ben, Lohmueller, Jason, Hostetter, Elizabeth,

- Cotsapas, Chris, ... Stewart, John. (2007). Genome-wide detection and characterization of positive selection in human populations. *Nature*, 449(7164), 913–918. <https://doi.org/10.1038/nature06250>
- Sankararaman, Sriram, Mallick, Swapan, Dannemann, Michael, Prüfer, Kay, Kelso, Janet, Pääbo, Svante, ... Reich, David. (2014). The genomic landscape of Neanderthal ancestry in present-day humans. *Nature*, 507(7492), 354–357. <https://doi.org/10.1038/nature12961>
- Saw, Woei Yuh, Tantoso, Erwin, Begum, Husna, Zhou, Lihan, Zou, Ruiyang, He, Cheng, ... Teo, Yik Ying. (2017). Establishing multiple omics baselines for three Southeast Asian populations in the Singapore Integrative Omics Study. *Nature Communications*, 8(1), 1–11. <https://doi.org/10.1038/s41467-017-00413-x>
- Somel, Mehmet, Creely, Hilliary, Franz, Henriette, Mueller, Uwe, Lachmann, Michael, Khaitovich, Philipp, & Pääbo, Svante. (2008). Human and Chimpanzee Gene Expression Differences Replicated in Mice Fed Different Diets. *PLoS ONE*, 3(1), e1504. <https://doi.org/10.1371/journal.pone.0001504>
- Southam, Andrew D., Weber, Ralf J. M., Engel, Jasper, Jones, Martin R., & Viant, Mark R. (2017). A complete workflow for high-resolution spectral-stitching nanoelectrospray direct-infusion mass-spectrometry-based metabolomics and lipidomics. *Nature Protocols*. <https://doi.org/10.1038/nprot.2016.156>
- Stranger, Barbara E., Montgomery, Stephen B., Dimas, Antigone S., Parts, Leopold, Stegle, Oliver, Ingle, Catherine E., ... Dermitzakis, Emmanouil T. (2012). Patterns of Cis Regulatory Variation in Diverse Human Populations. *PLoS Genetics*, 8(4),

e1002639. <https://doi.org/10.1371/journal.pgen.1002639>

Sud, M., Fahy, E., Cotter, D., Brown, A., Dennis, E. A., Glass, C. K., ... Subramaniam,

S. (2007). LMSD: LIPID MAPS structure database. *Nucleic Acids Research*,
35(Database), D527–D532. <https://doi.org/10.1093/nar/gkl838>

Sugimoto, Masahiro, Sakagami, Hiroshi, Yokote, Yoshiko, Onuma, Hiromi, Kaneko,

Miku, Mori, Masayo, ... Tomita, Masaru. (2012). Non-targeted metabolite profiling
in activated macrophage secretion. *Metabolomics*, 8(4), 624–633.

<https://doi.org/10.1007/s11306-011-0353-9>

Tang, Yuchun, Zhao, Lu, Lou, Yunxia, Shi, Yonggang, Fang, Rui, Lin, Xiangtao, ...

Toga, Arthur. (2018). Brain structure differences between Chinese and Caucasian
cohorts: A comprehensive morphometry study. *Human Brain Mapping*, 39(5),
2147–2155. <https://doi.org/10.1002/hbm.23994>

Tishkoff, S. A., Reed, F. A., Friedlaender, F. R., Ehret, C., Ranciaro, A., Froment, A., ...

Williams, S. M. (2009). The Genetic Structure and History of Africans and African
Americans. *Science*, 324(5930), 1035–1044.

<https://doi.org/10.1126/science.1172257>

Tishkoff, S. A., & Williams, S. M. (2002). Genetic analysis of African populations:

human evolution and complex disease. *Nature Reviews Genetics*, 3(8), 611–621.

<https://doi.org/10.1038/nrg865>

Tomasello, M. (2019). Becoming human: A theory of ontogeny. In *Belknqp*.

Turakulov, Rust, & Easteal, Simon. (2003). Number of SNPS loci needed to detect
population structure. *Human Heredity*, 55(1), 37–45.

<https://doi.org/10.1159/000071808>

Van Laer, Bart, Kapp, Ulrike, Soler-Lopez, Montserrat, Moczulska, Kaja, Pääbo, Svante, Leonard, Gordon, & Mueller-Dieckmann, Christoph. (2018). Molecular comparison of Neanderthal and Modern Human adenylosuccinate lyase. *Scientific Reports*, 8(1), 18008. <https://doi.org/10.1038/s41598-018-36195-5>

Vance, J. E., & Vance, Dennis E. (2008). Biochemistry Of Lipids, Lipoproteins And Membranes. In *Biochemistry of Lipids, Lipoproteins and Membranes*. <https://doi.org/10.1016/B978-0-444-53219-0.X5001-6>

Varki, A. (2005). Comparing the human and chimpanzee genomes: Searching for needles in a haystack. *Genome Research*, 15(12), 1746–1758. <https://doi.org/10.1101/gr.3737405>

White, Tim D., Asfaw, Berhane, DeGusta, David, Gilbert, Henry, Richards, Gary D., Suwa, Gen, & Clark Howell, F. (2003). Pleistocene Homo sapiens from Middle Awash, Ethiopia. *Nature*, 423(6941), 742–747. <https://doi.org/10.1038/nature01669>

Williams, AL, Jacobs, SB, Moreno-Macías, H., Huerta-Chagoya, A., Churchhouse, C., Márquez-Luna, C., ... Altshuler, D. (2014). Sequence variants in SLC16A11 are a common risk factor for type 2 diabetes in Mexico. *Nature*, 506(7486), 97–101. <https://doi.org/10.1038/nature12828>

Wishart, David S., Feunang, Yannick Djoumbou, Marcu, Ana, Guo, An Chi, Liang, Kevin, Vázquez-Fresno, Rosa, ... Scalbert, Augustin. (2018). HMDB 4.0: The human metabolome database for 2018. *Nucleic Acids Research*. <https://doi.org/10.1093/nar/gkx1089>

- Wright Willis, Allison, Evanoff, Bradley A., Lian, Min, Criswell, Susan R., & Racette, Brad A. (2010). Geographic and ethnic variation in Parkinson disease: A population-based study of us medicare beneficiaries. *Neuroepidemiology*, *34*(3), 143–151.
<https://doi.org/10.1159/000275491>
- Xie, Liang, Wu, Wei, Chen, Jin, Tu, Jianglong, Zhou, Jun, Qi, Xueliang, & Yin, Xiaoping. (2017). Cholesterol Levels and Hemorrhagic Stroke Risk in East Asian Versus Non-East Asian Populations. *The Neurologist*, *22*(4), 107–115.
<https://doi.org/10.1097/NRL.0000000000000126>
- Yang, Kui, Cheng, Hua, Gross, Richard W., & Han, Xianlin. (2009). Automated Lipid Identification and Quantification by Multidimensional Mass Spectrometry-Based Shotgun Lipidomics. *Analytical Chemistry*, *81*(11), 4356–4368.
<https://doi.org/10.1021/ac900241u>
- Yang, Kui, & Han, Xianlin. (2016). Lipidomics: Techniques, Applications, and Outcomes Related to Biomedical Sciences. *Trends in Biochemical Sciences*, *41*(11), 954–969. <https://doi.org/10.1016/j.tibs.2016.08.010>
- Yu, Jin-Tai, Tan, Lan, & Hardy, John. (2014). Apolipoprotein E in Alzheimer's Disease: An Update. *Annual Review of Neuroscience*, *37*(1), 79–100.
<https://doi.org/10.1146/annurev-neuro-071013-014300>
- Zabrecky, James R., & Raftery, Michael A. (1985). The role of lipids in the function of the acetylcholine receptor. *Journal of Receptors and Signal Transduction*.
<https://doi.org/10.3109/10799898509041890>
- Zheng, Fanfan, Yan, Li, Yang, Zhenchun, Zhong, Baoliang, & Xie, Wuxiang. (2018).

HbA1c, diabetes and cognitive decline: the English Longitudinal Study of Ageing.

Diabetologia, 61(4), 839–848. <https://doi.org/10.1007/s00125-017-4541-7>

Zheng, Xiaojiao, Chen, Tianlu, Zhao, Aihua, Wang, Xiaoyan, Xie, Guoxiang, Huang,

Fengjie, ... Jia, Wei. (2016). The brain metabolome of male rats across the lifespan.

Scientific Reports. <https://doi.org/10.1038/srep24125>

Zulfiqar, Maria, Lin, Doris D. M., Van der Graaf, Marinette, Barker, Peter B., Fahrner,

Jill A., Marie, Sandrine, ... Maegawa, Gustavo H. B. (2013). Novel proton MR

spectroscopy findings in adenylosuccinate lyase deficiency. *Journal of Magnetic*

Resonance Imaging, 37(4), 974–980. <https://doi.org/10.1002/jmri.23852>