



Skolkovo Institute of Science and Technology

SKOLKOVO INSTITUTE OF SCIENCE AND TECHNOLOGY

COMPARATIVE ANALYSIS OF HUMAN
BRAIN BASED ON
MASS SPECTROMETRY DATA

Doctoral Thesis

by

ILIA KUROCHKIN

DOCTORAL PROGRAM IN LIFE SCIENCE

Supervisor

Professor Philipp Khaitovich

Moscow - 2018

© Ilia Kurochkin - 2018

Abstract

Even though metabolites and lipids play critical role in brain function, little is known about the brain's metabolome and lipidome composition, and their differences between the brain and other tissues. The alteration of brain metabolism in cognitive disorders and the differences between the human brain metabolism and the brain metabolism of other primate species are similarly unexplored. Neurocognitive disorders are a complex group of diseases characterized by the disruption of evolutionarily novel cognitive skills, including communicative and social abilities. Metabolite and lipidome measurements in urine and blood plasma indicate disruption of particular pathways in neurocognitive disorders, including autism spectrum disorder (ASD) and schizophrenia (SZ). Yet, the relationship of these findings to brain remained poorly explored.

In this work, we investigated the alteration in metabolome and lipidome composition in neurocognitive disorders, including ASD and SZ, comparing prefrontal cortex from cognitively affected patients and healthy controls. We found a substantial fraction of metabolites and lipids with concentrations significantly altered in diseases. Some of the changes could be connected to previous observations made in urine and blood. Furthermore, we found an excess of human-specific metabolite concentration differences in several pathways altered in ASD. The human-specific metabolites were identified in comparison between humans and chimpanzees and macaques. To extend this evolutionary comparison, we analyzed lipidome evolution in six tissues of 32 species representing primates, rodents, and bats. We showed that many of the uniquely human lipidome features localize in the brain cortex and cluster in specific pathways implicated in ASD, SZ and other neurocognitive disorders.

Along with several previous studies, our results support the idea that evolution of cognitive abilities in human was accompanied by adaptive changes in brain metabolism.

Acknowledgements

I wish to express my very great appreciation to my advisor Philipp Khaitovich and to Khaitovich's lab members: Ekaterina Khrameeva, Vita Stepanova, Anna Tkachev, Pavel Mazin, Anna Vanyushkina, Aleksandra Mitina, Dmitry Zubkov.

Contents

Abstract	ii
Acknowledgements	iii
List of Figures	vii
Abbreviations	ix
1 Introduction	1
1.1 Significance of the work	1
1.2 Project Objectives	2
1.3 Novelty and Practical Use	2
1.4 Personal Contribution	3
1.5 Publications	3
2 Review of Literature	5
2.1 Definition of complex phenotypes	5
2.2 Genetic architecture of autism spectrum disorder	7
2.3 Metabolome and Lipidome levels	9
2.4 Metabolites of the cell and their involvement in neurological disorder	9
2.4.1 Overview of the metabolome	9
2.4.2 Metabolome alterations in autism	11
2.5 Lipids of the cell and their involvement in neurological disorder	15
2.5.1 Diversity of membrane lipid chemical structures	16
2.5.2 Lipid functionality	18
2.5.3 Lipidome alterations in autism	20
2.5.4 Lipidome alterations in schizophrenia	22
2.6 Evolutionary models at different molecular levels	23
2.6.1 Evolutionary studies at the level of genome and transcriptome	23
2.6.2 Evolutionary studies at the level of metabolome and lipidome	25

3	Metabolome signature of autism in the human prefrontal cortex	29
3.1	Materials and Methods	29
3.1.1	Samples	29
3.1.2	Sample preparation	30
3.1.3	Mass spectrometry analysis	31
3.1.4	Mass spectrometry data processing	32
3.1.5	Statistical analysis	33
3.2	Results	37
3.2.1	Metabolic data description	37
3.2.2	Metabolic changes in autism	37
3.2.3	Metabolic predictors of autism	39
3.2.4	Gene expression associated with metabolic changes in autism	41
3.2.5	Evolution of metabolic changes in autism	41
3.3	Discussion	44
4	Lipidome alterations in human prefrontal cortex in cognitive disorders	47
4.1	Materials and Methods	47
4.1.1	Samples	47
4.1.2	MS sample preparation and measurements	48
4.1.3	Lipidome data preprocessing	49
4.1.4	Lipid annotation and enrichment analysis	50
4.1.5	Disorder-associated lipids identification	50
4.1.6	Quantification of disorder effect on lipid concentration variance	51
4.2	Results	51
4.2.1	The PFC lipidome composition is altered in common cognitive disorders	51
4.2.2	Membrane fluidity alterations inferred based on lipidome composition	55
4.3	Discussion	56
5	Lipidome evolution in mammalian tissues	59
5.1	Materials and Methods	59
5.1.1	Samples	59
5.1.2	Normalization of concentrations	60
5.1.3	Peak annotation	60
5.1.4	Identification of lipids and genes that follow phylogeny	61
5.1.5	Identification of protein coding sequences that follow phylogeny	63

5.1.6	Analysis of lipid concentration differences within and between species	63
5.1.7	Identification of species-specific lipids	63
5.1.8	Comparison with published data set	64
5.1.9	Human-specific gene expression changes	65
5.2	Results	65
5.2.1	Lipidome data description	65
5.2.2	Lipidome differences and phylogenetic distances among species	66
5.2.3	Species-specific lipidome differences	68
5.3	Discussion	72
6	Conclusions	75
	Bibliography	78

List of Figures

2.1	Genetic architecture of ASD [de la Torre-Ubieta et al., 2016]	8
2.2	Pathways enrichment analysis in urine of ASD patients	12
2.3	Diversity of lipids composing the membrane [Harayama and Riezman, 2018]	17
2.4	Membrane properties affected by lipid composition [Harayama and Riezman, 2018]	18
2.5	Lipids mediate their function through lipid-protein interactions [Harayama and Riezman, 2018]	19
3.1	Concentration profiles of subset of metabolites affected by postmortem delay	33
3.2	Area under the receiver operating characteristic curve (ROC AUC)	35
3.3	Identification of metabolic differences in the prefrontal cortex of autism patients	38
3.4	Characterization of metabolic differences identified in the prefrontal cortex of autism patients	40
3.5	Evolution of metabolic differences identified in the prefrontal cortex of autism patients	42
4.1	Disorder-associated lipidome changes in the human PFC [Yu et al., 2018]	52
4.2	Lipidome and transcriptome alterations in the three cognitive diseases [Yu et al., 2018]	53
4.3	Autism-associated lipidome changes in DS2 [Yu et al., 2018]	54
4.4	Predicted membrane fluidity levels in disorders [Yu et al., 2018]	56
5.1	Dependence between R^2 values and R^2 permutation p-values [Khrameeva et al., 2018]	62
5.2	Lipidome data overview [Khrameeva et al., 2018]	66
5.3	Relationship between lipid concentrations and phylogenetic distances [Khrameeva et al., 2018]	67
5.4	Species-specific lipid concentration differences [Khrameeva et al., 2018]	69

5.5	Characterization of lipids showing human-specific concentration levels in cortex [Khrameeva et al., 2018]	71
5.6	Evolution at lipidome, transcriptome and genome levels [Khrameeva et al., 2018]	72

Abbreviations

2-AG	2-arachydonoyl glycerol
ANCOVA	Analysis of covariance
ANOVA	Analysis of variance
AA	Arachidonic acid
AEA	Arachidonylethanolamide
AUC	Area under the curve
ASD	Autism spectrum disorder
BH correction	Benjamini-Hochberg correction
BCFA	Branched-chain fatty acids
CE	Cholesterol ester
CNV	Copy-number variation
DNA	Deoxyribonucleic acid
DG	Diglycerides
DHA	Docosahexaenoic acid
DS	Down syndrome
EPA	Eicosapentaenoic acid
eCB	Endocannabinoid
FPKM	Fragments per kilobase million
FC	Free cholesterol
FA	Free fatty acids
GC-MS	Gas chromatography coupled with mass spectrometry
GWAS	Genome-wide association studies
GP	Glycerophospholipid
HDL-C	High-density lipoprotein cholesterol
LC-MS	Liquid chromatography coupled with mass spectrometry
LDL-C	Low-density lipoprotein cholesterol
LyPC	Lysophosphatidylcholine
MG	Monoglycerides
MUFA	Monounsaturated fatty acids
MDS	Multidimensional scaling
NMR	Nuclear magnetic resonance
OCFA	Odd-chain fatty acids
OR	Odds ratio
OLS	Ordinary least squares

OU model/process	Ornstein-Uhlenbeck model/process
OPLS-DA	Orthogonal projections to latent structures discriminant analysis
PLS	Partial least squares regression
PC	Phosphatidylcholine
PE	Phosphatidylethanolamine
PG	Phosphatidylglycerol
PI	Phosphatidylinositol
PS	Phosphatidylserine
PUFA	Polyunsaturated fatty acids
PMD	Post-mortem delay
PMI	Post-mortem interval
PFC	Prefrontal cortex
PGE2	Prostaglandin E2
PG-G	Prostaglandin H2
ROC	Receiver operating characteristic
RT	Retention time
RNA	Ribonucleic acid
SFA	Saturated fatty acids
SZ	Schizophrenia
SNP	Single nucleotide polymorphisms
SNV	Single nucleotide variant
SM	Sphingomyelins
SVM	Support vector machine
TCA cycle	The tricarboxylic acid cycle
TC	Total cholesterol
TFA	Trans fatty acids
TG	Triglycerides

Chapter 1

Introduction

1.1 Significance of the work

Autism, schizophrenia, and other neurocognitive disorders involve disruption of cognitive functions related to socialization and communication, the key behavioral traits that separate humans from other species. Multiple studies focusing on identification of the mechanisms of neurocognitive disorders mainly investigated the genetic differences between disease and control groups. These studies showed that autism, schizophrenia, and other neurocognitive disorders could be linked to a large number of genetic variants, each having a small effect. In the absence of clear genetic determinants common to the disease, several laboratories searched for potential biochemical landmarks of autism and schizophrenia by monitoring concentrations of low weight molecular compounds (metabolites) in urine and blood plasma. While these studies yielded a number of metabolite concentration differences between disease and control groups, investigation of metabolic differences in brain samples was still lacking. The studies presented in my thesis thus represent the first large-scale analyses of metabolome and lipidome differences in common neurocognitive disorders, autism, and schizophrenia. We further investigated these disease-associated differences in the context of evolutionary diversity in brain metabolism in primates and mammals, including identification of metabolic features unique to humans. In these studies, we identified metabolic pathway altered in brains of autism and schizophrenia patients. We further determined the overlap of these disease-associated pathways with human-specific differences identified in our analysis of brain metabolome and lipidome of humans, chimpanzees, and macaques, as well as broader evolutionary analysis of 32 mammalian species. The identified pathways represent promising candidates for in-depth investigation of the disease mechanisms and for further studies targeting the advanced social and learning abilities unique to humans. We further demon-

strated that many of metabolome and lipidome alterations in prefrontal cortex detected in autism, both at the pathway level and at the level of individual metabolites, coincide with the differences reported in urine and blood samples. This finding underscores the potential of metabolic studies conducted in non-nervous tissues.

1.2 Project Objectives

The goal of this work is to assess the metabolome and lipidome alteration in cognitive disorders and compare those differences with functional metabolic differences corresponding to evolutionarily novel, human-specific processes.

Several specific aims were addressed for this purpose:

- Investigate metabolome alteration in prefrontal cortex in autism patients and identification of human-specific metabolome changes.
- Study lipidome alteration in prefrontal cortex in several neurocognitive disorders: autism, schizophrenia, and Down syndrome.
- Examine lipidome evolution in brain and non-neuronal tissues in 32 mammalian species representing three distinct clades: primates, rodents, and bats.

1.3 Novelty and Practical Use

AA comprehensive study of metabolome and lipidome in ASD was performed. This is the first large-scale analysis, which was done in the prefrontal cortex, the brain region implicated in high-level cognitive functions and known to be affected in common neurocognitive disorders. We identified pathways that were not previously linked to ASD, which include pyrimidine metabolism, beta-alanine metabolism. Additionally, we demonstrated that many of the metabolome and lipidome alteration found in brain coincide with metabolic pathways identified in studies carried out in urine and blood.

We were the first who addressed the general mode of lipidome evolution at a large scale by applying standard evolutionary analysis algorithms to a large lipidome dataset containing data for 18,000 lipids measured in 6 tissues of 32 mammalian species. We demonstrated that lipid concentration levels evolve differently compared to genome sequences or gene expression levels and their evolutionary mode cannot be described adequately by the evolutionary models developed for genome and transcriptome evolution.

The practical outcome of this work is the potential of using non-nervous tissues for identification of key metabolome alteration in cognitive diseases. Our work indicates that metabolites and lipids evolve differently from genes and transcripts, and therefore, might yield very different results in studies of intra population variation, including disease susceptibility modeling.

1.4 Personal Contribution

Most of the bioinformatics analyses presented in this thesis was performed by the author. Specifically, all analysis, except identification of metabolic predictors of autism using logistic regression, within the chapter describing the identification and characterization of metabolome signature of autism in the human prefrontal cortex was solely performed by the author. In the chapter describing lipidome alteration in human prefrontal cortex in three cognitive disorders, the author analyzed lipidome alteration in autism and performed validation of results based on an independent dataset. In the lipidome evolution chapter, the author participated in all analysis steps with particular focus on analyses dedicated to identification and characterization of lipidome differences, which followed phylogeny, as well as analyses of species-specific lipidome differences, including the ones based on the use of Ornstein-Uhlenbeck model. The author also performed parallel analysis of molecular traits, which follow phylogeny, at the levels of transcriptome and genome.

1.5 Publications

The thesis-related results were presented at a scientific conference and published in international peer-reviewed journals.

List of publications:

1. **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. (Under review in *Communications Biology*) (Chapter 3)
2. Yu Q*, He Z*, Zubkov D*, Huang S, **Kurochkin I**, Yang X, Halene T, Willmitzer L, Giavalisco P, Akbarian S, Khaitovich P. Lipidome alterations in human prefrontal cortex during development, aging, and cognitive disorders. *Mol Psychiatry*. 2018 Aug 8. doi: 10.1038/s41380-018-0200-8. (Chapter 4)

3. Khrameeva E*, **Kurochkin I***, Bozek K, Giavalisco P, Khaitovich P. Lipidome Evolution in Mammalian Tissues, *Mol Biol Evol.* 2018 May 11. doi: 10.1093/molbev/msy097. (Chapter 5)

List of conferences:

1. SMBE 2018. **Kurochkin I***, Khrameeva E, Tkachev A, Khaitovich P. Molecular signatures of autism in prefrontal cortex. July 8 - 12, 2018, Yokohama, Japan. (Chapter 3)
2. ITAS 2016. Khrameeva E*, **Kurochkin I.** Lipidome determinants of lifespan in mammals. September 25 - 30, Repino, Russia. (Chapter 5)
3. Metabolomics 2016. **Kurochkin I***, Khaitovich P. Metabolic changes in prefrontal cortex of humans, chimpanzees and macaques during postnatal development. June 27 - 30, 2016, Dublin, Ireland. (Chapter 3)

Chapter 2

Review of Literature

2.1 Definition of complex phenotypes

From the point of view of the underlying genetic background, human phenotypes are often classified as either simple or complex. In contrast to simple phenotypes, which can be explained by the action of a single gene, complex phenotypes, including complex diseases, are controlled by multiple genes and include interactions between genes and environmental factors. We define complex traits as those that do not follow readily predictable patterns of inheritance (Mendelian Inheritance patterns) and commonly exhibit a large variety of phenotypes. Usually, loci associated with complex traits are spread across the genome and include genes showing no apparent connection to phenotype [Boyle et al., 2017].

Due to the development of sequencing technology, it became possible to sequence thousands of individuals, giving rise to genome-wide association studies (GWAS). The multiple GWAS conducted to date for a number of complex traits, including cognitive diseases, has provided a detailed understanding of the genetic basis of these traits. It was demonstrated that complex traits are mainly driven by noncoding variants that might affect gene regulation, in contrast to Mendelian traits, which are primarily caused by protein-coding changes [Boyle et al., 2017]. In addition to genomic differences, complex traits can be caused by changes in gene expression levels, splicing, histone modifications and other levels. Moreover, combining information of genomic variations with different molecular levels gave rise to series of quantitative trait loci (QTLs) studies, in particular, expression QTL (eQTL), metabolic QTL (mQTL), splicing QTL (sQTL). The primary goal of such studies is to find genetic variants which explain variation at different molecular levels, in the case of eQTLs genetic variants which explain the variation of gene expression levels. It was shown that sQTL and

eQTL could modulate phenotypic traits, such as rheumatoid arthritis, multiple sclerosis, Alzheimer’s disease, schizophrenia, height, and body mass index [Li et al., 2016].

An example of a successful investigation of relatively simple genetic determinants of a molecular trait are studies of cholesterol and lipoprotein levels. Multiple studies investigated levels of cholesterol concentration in blood, as alteration of cholesterol concentrations could lead to cardiovascular disease and other health problems [Cohen, 2004, Willer et al., 2008, the ENGAGE Consortium et al., 2009]. In one study, the authors investigated whether rare sequence variations contribute to low plasma levels of HDL-C. For this purpose, they sequenced three candidate genes (*ABCA1*, *APOA1*, and *LCAT*) in Dallas Heart Study cohorts including 128 persons with low and 128 persons with high plasma HDL-C levels from each of four gender and race groups (black men, white men, black women, and white women). They found that nonsynonymous sequence variants with significant phenotypic effects contribute to low plasma HDL-C levels in all four gender and race groups. Additionally, they obtained similar results based on an independent Canadians cohort of 155 persons with low HDL-C and 108 persons with high HDL-C [Cohen, 2004]. In another study, the authors identified loci affecting serum levels of TC, LDL-C, HDL-C, and TG across different European populations and showed that genetic risk scores could explain up to 4.8% of the variation in lipid concentrations. A similar percentage of lipid concentration variation can be explained by the body mass index (BMI), which is commonly used as a traditional risk factor for cholesterol levels [the ENGAGE Consortium et al., 2009].

In the past decade, numerous studies aimed to identify variants with pleiotropic effects on different complex traits. These studies yielded multiple genetic variants associated with a given trait with varying strength, commonly explaining no more than one percent of the phenotype. Furthermore, some of the variants overlapped among multiple traits [Pickrell et al., 2016, ReproGen Consortium et al., 2015, Cotsapas et al., 2011]. For example, in a recent study, the authors assessed pleiotropic genetic effect for 42 human traits divided into five distinct groups: neurological phenotypes, anthropometric and social traits, immune-related traits, metabolic phenotypes and hematopoietic traits [Pickrell et al., 2016]. They identified multiple loci that are associated with multiple traits and ranked those loci according to the number of phenotypes shared between these particular genetic variants. The top locus based on that ranking was a nonsynonymous variant in *SH2B3* gene associated with a number of autoimmune diseases, lipid traits, heart disease, and red blood cell traits. Besides that, the top loci included a nonsynonymous variant in *SLC39A8* gene associated with schizophrenia, Parkinson disease, high-density lipoprotein cholesterol, and height and a set of variants near

APOE gene associated with Alzheimer’s disease, low-density lipoprotein cholesterol, total cholesterol, triglycerides and high-density lipoprotein cholesterol.

One of the main results of GWAS is what they do not identify genetic variants with a strong association to the trait. By contrast, most of the identified variants explained less than one percent of the phenotypic variation. Thus, most of the complex traits cannot be explained by common variants identified thus far. It is possible that the missing genetic variance is coming from yet unidentified common genetic variants with small effect sizes. Another possible explanation could be that missing variation can be explained by rare and novel variants, which are not present in SNPs database; as a result, GWAS studies are not able to capture their contribution to complex traits [Frazer et al., 2009]. Other explanations were offered as well [Frazer et al., 2009, Cohen, 2004, Shao et al., 2008]. Most importantly, GWAS did not deliver reliable genetic markers for most of the complex traits, including common disorders.

2.2 Genetic architecture of autism spectrum disorder

Autism spectrum disorder (ASD) is a complex group of diseases characterized by the developmental disruption of complex cognitive traits, particularly the ones underlying communication and socialization. ASD is a heterogeneous condition characterized by a broad spectrum of symptoms’ severity and can occur with frequent comorbidities.

Genetic architecture of ASD is characterized by many factors including mode of inheritance (for example, autosomal inheritance, X-linked inheritance, and *de novo* variation), type of variation (for example, copy number variation, translocations, indels, and single-nucleotide variation), mode of action (additive, dominant, recessive or hemizygous) and frequency in the population [de la Torre-Ubieta et al., 2016, De Rubeis and Buxbaum, 2015] (Figure 2.1).

Two major mechanistic models were proposed to describe the nature of ASD: polygenic and major gene model [de la Torre-Ubieta et al., 2016]. The polygenic model claims that many inherited variants cause ASD, each having small effects, together contributing to ASD. In contrast, major gene model claims that one or several genetic variants, each having a substantial risk, can cause ASD. Additionally, there is a hypothesis that one dominant mutation in a gene might be enough to provoke ASD. Major gene model and polygenic model are not mutually exclusive, as there is support in the literature of both models.

Numerous genetic studies indicate that most ASD cases are likely to be caused by a large number of genetic variants each having a small effect, including rare events of chromosomal rearrangements, duplications, and deletions, single-nucleotide variations, and indels. Such

risk variants of ASD could be both inherited and have risen *de novo* in the process of meiosis during gametogenesis. Autosomal *de novo* loss-of-function SNV has been associated with the highest risk of ASD development, while transmitted variants that disrupt proteins have a lower risk [The DDD Study et al., 2014]. Advanced paternal age is believed to increase the number of *de novo* SNV and CNV, while advanced maternal age is a known risk factor for chromosomal aberrations [Sandin et al., 2016, De Rubeis and Buxbaum, 2015].

At the same time, there are rare ASD cases caused by one or several genetic mutations. The most well-known examples include *CHD8*, *DYRK1A*, *SCN2A*, *ARID1B*, *ADNP*, *ANK2*, *TBR1*.

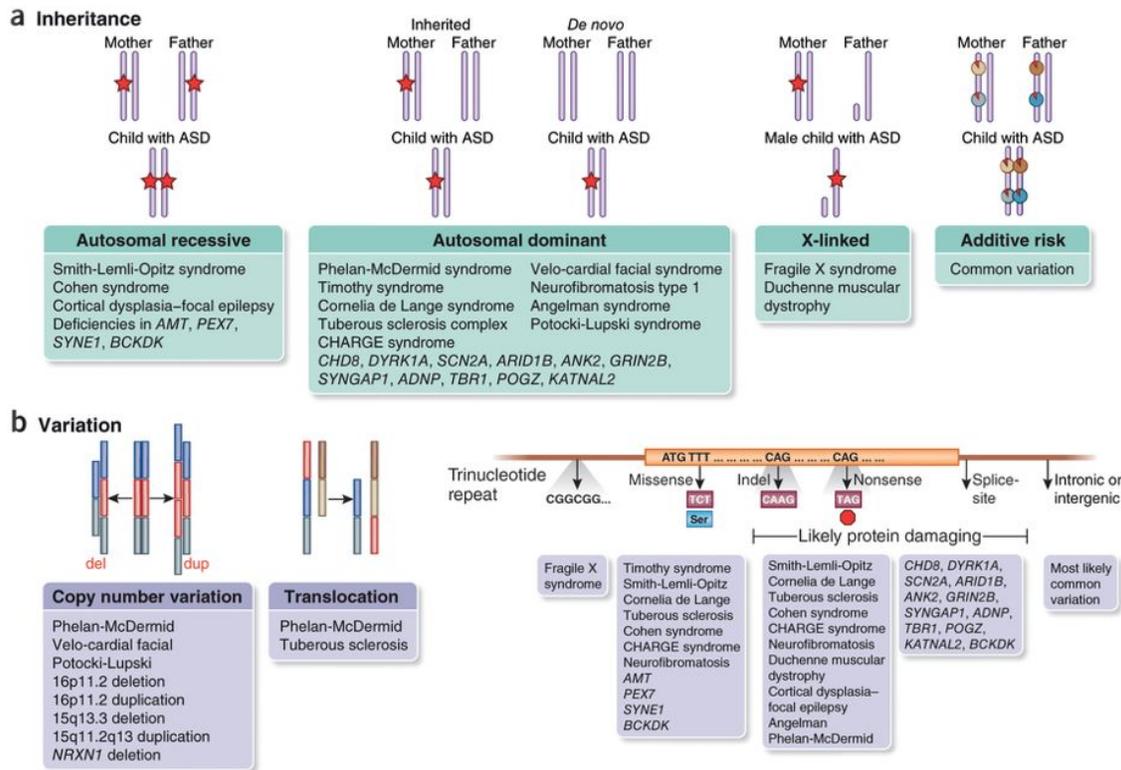


Figure 2.1: Genetic architecture of ASD (reproduced with permission from Nature Medicine [de la Torre-Ubieta et al., 2016]).

At present, there are ≈ 1200 genes associated with ASD risk [Sanders et al., 2012, The DDD Study et al., 2014]. Disruption in these genes was associated with multiple human disorders. This agrees with the fact that ASD occurs with frequent comorbidities. For example, *ADNP* gene that is involved in chromatin remodeling, autophagy, and microtubule dynamics was shown to be associated with deficiencies in spatial learning, object recognition, social recognition, and social memory. Besides, knockout of *ADNP* gene in mouse causes

embryonic lethality. *ANK2* gene has been associated with ASD and cardiac arrhythmia, while knockout of *ANK2* gene in mouse leads to postnatal lethality [De Rubeis and Buxbaum, 2015].

2.3 Metabolome and Lipidome levels

In addition to the genome, it is also possible to assess the differences characteristic of ASD patients at other levels of molecular phenotype: metabolome and lipidome. The terms metabolome and lipidome refer to the sum of small molecular weight molecules contained within cells and tissues. The lipidome represents mostly non-polar compounds, while the metabolome commonly refers to the polar and amphiphilic ones. In the context of this thesis, we define metabolites as the hydrophilic (polar) fraction of small biological molecules with a molecular weight <1500Da. In contrast, lipids represent the hydrophobic fraction of small biological molecules with a molecular weight <1500Da. This classification is introduced due to the methodology of metabolites/lipids extraction.

2.4 Metabolites of the cell and their involvement in neurological disorder

2.4.1 Overview of the metabolome

Hydrophilic (polar) metabolites are omnipresent within the human organism and include all essential compound classes, such as carbohydrates, amino acids, nucleotides, vitamins, and coenzymes. According to widely used classification, all the polar metabolites could be divided into primary and secondary metabolites.

Primary metabolites are molecules of small molecular weight involved in crucial biological processes, such as growth, development, energy consumption, cell division, and proliferation and typically involved in many different metabolic pathways [James, 2017]. Secondary metabolites are non-essential for primary organism activity, but ensure the survival advantages by enhancing the environmental interactions, stress response and modulating the metabolism [Breitling et al., 2013].

One of the common primary metabolites that animal tissues utilize as their universal energy source is glucose. Glucose is metabolized in a variety of pathways, *e.g.*, glycolysis and gluconeogenesis, pentose and glucuronate interconversions, starch and sucrose metabolism and others. Animal food is a main source of glucose. However, kidney and liver have ma-

chinery for *de novo* synthesis of glucose from non-glucose precursors and release glucose into the circulation by the breakdown of glycogen [Stumvoll, 1998]. In addition to its energetic function, glucose metabolism also provides the source for the biosynthesis of neurotransmitters [Mergenthaler et al., 2013], amino acids (e.g. synthesis of alanine from pyruvate) [Garber et al., 1976] and the agents that are responsible for blood flow regulation [Lombard, 2006].

Another example of a primary metabolite with plenty of functions is adenosine triphosphate (ATP). A crucial building component of nucleic acid molecules and the main cellular energy source, ATP is also involved in extracellular communication as a mediator of both immunologic responses and stem cell functions [Rossi et al., 2012].

Secondary metabolites, as primary ones, have varied chemical structures. Despite the diversity of secondary metabolites produced by microorganisms and plants, they are often synthesized from a small amount of metabolized products, e.g., acetyl-CoA, amino acids, and intermediates of the shikimic acid pathway and the tricarboxylic acid cycle [Demain and Fang, 2000]. Antibiotics are a typical example of secondary metabolites that are commonly used in medicine, agriculture, and basic research [Demain and Fang, 2000].

Primary and secondary metabolites are tightly bound in metabolic pathways and influence each other via feedback regulation, induction, and catabolite repression [Demain and Fang, 2000]. There are no uniform structures or chemical units specific to polar metabolites except for their hydrophilicity, and their functions range from storage and transmission of genetic information to cell signaling.

The average concentration of each metabolite varies from tissue to tissue and depends on intracellular cell localization. For example, the most abundant polar metabolites found in serum are D-glucose, urea, ATP, glyceraldehyde, L-lactic acid and fructosamine, L-glutamine, L-alanine, methanol, glycine, L-lysine, uric acid, and (R)-3-hydroxybutyric acid [Psychogios et al., 2011]. Interestingly, the concentration of glucose in the human brain, which utilizes it as a sole energy source, is about 1 mM while the serum the glucose level is 5 mM.

A tissue metabolic profile is heavily influenced by its function. For instance, since one of the major kidney functions results in urine production, urea concentration in urine may be a 100-fold higher than in plasma and extracellular fluids in humans [Bankir and Yang, 2012]. Nevertheless, changes in the concentration of specific metabolites could indicate stress or disease condition. For instance, concentrations of compounds involved in purine metabolism decrease in blood serum of bipolar patients [Gubert et al., 2016]. Similarly, the concentration of γ -aminobutyric acid in cerebrospinal fluids could indicate the severity of Parkinson's disease [Araki et al., 2009], while hypoxanthine and inosine concentrations are known to be significantly elevated in serum after kidney reperfusion [Domański et al., 2006].

2.4.2 Metabolome alterations in autism

In the absence of clear genetic determinants, previous studies searched for potential biochemical landmarks of ASD by monitoring concentrations of low molecular weight metabolites in urine, blood plasma, and brain.

To date, metabolite concentration differences in urine of ASD patients were mostly studied using systematic metabolomics profiling techniques, yielding results for 20 to 622 metabolites measured in 14 to 48 ASD patients and comparable control groups [Yap et al., 2010, Ming et al., 2012, Emond et al., 2013, Mavel et al., 2013, Cozzolino et al., 2014, Nadal-Desbarats et al., 2014, Kałużna-Czaplińska et al., 2014, Noto et al., 2014, Diémé et al., 2015, Gevi et al., 2016, Lussu et al., 2017, Bitar et al., 2018]. Summary of the pathway enrichment analysis based on those studies is shown in Figure 2.2.

The pioneering study conducted using proton-NMR and involving 39 ASD patients, 28 healthy siblings, and 34 healthy volunteers demonstrated alterations of the amino acid metabolism and the tryptophan and nicotinic acid metabolic pathways [Yap et al., 2010]. Authors have built OPLS-DA model that distinguished each of three groups in pairwise comparisons. They demonstrated that several metabolites separate ASD group from controls. Among them, acetate, dimethylamine (DMA), N-acetyl glycoprotein fragments (NAG), succinate, taurine, N-methyl nicotinic acid (NMNA), N-methyl nicotinamide (NMND) and N-methyl-2-pyridone-5-carboxamide (2PY) were increased in ASD, while glutamate, hippurate, and phenylacetylglutamine (PAG) were decreased in ASD compared to controls. Comparison between healthy siblings and other two groups (ASD and controls) using OPLS-DA showed no differences. To validate metabolites obtained between ASD patients and controls by OPLS-DA, they used Mann-Whitney test. Only NMNA and NMND remained significantly different, and there were additional significant differences in other pairwise comparisons between groups.

Another study identified differences in amino acid metabolism, as well as metabolic signatures of oxidative stress, in the urine of 48 children with ASD and 53 age-matched controls using a combination of liquid chromatography and gas chromatography coupled to mass spectrometry (LC-MS and GC-MS) [Ming et al., 2012]. The authors detected 391 metabolites with known chemical structures in the urine samples. Among those metabolites, the concentrations of 82 of them were significantly altered between the ASD and the control groups (53 were increased in ASD, while 29 were decreased) based on t-test. These 82 metabolites

	Ming et al., 2012	Bitar et al., 2018	Mavel et al., 2013	Lussu et al., 2017	Yap et al., 2010	Nadal-Desbarats et al., 2014	Kauna-Czaplińska et al., 2014	Gevi et al., 2016	Dim et al., 2015	Noto et al., 2014	Emond et al., 2013	Cozzolino et al., 2014
Aminoacyl-tRNA biosynthesis												
Phenylalanine metabolism												
Glycine, serine and threonine metabolism												
Nitrogen metabolism												
Propanoate metabolism												
Thiamine metabolism												
Citrate cycle (TCA cycle)												
Glutathione metabolism												
Glyoxylate and dicarboxylate metabolism												
Tryptophan metabolism												
Alanine, aspartate and glutamate metabolism												
Butanoate metabolism												
Cyanoamino acid metabolism												
Histidine metabolism												
Lysine degradation												
Methane metabolism												
Purine metabolism												
Tyrosine metabolism												
Arginine and proline metabolism												
Phenylalanine, tyrosine and tryptophan biosynthesis												
Porphyrin and chlorophyll metabolism												
Primary bile acid biosynthesis												
beta-Alanine metabolism												
Cysteine and methionine metabolism												
Pantothenate and CoA biosynthesis												
Pyrimidine metabolism												
Taurine and hypotaurine metabolism												
Ubiquinone and other terpenoid-quinone biosynthesis												
Ascorbate and aldarate metabolism												
Glycolysis or Gluconeogenesis												
Inositol phosphate metabolism												
Nicotinate and nicotinamide metabolism												
Pentose and glucuronate interconversions												
Starch and sucrose metabolism												
Valine, leucine and isoleucine biosynthesis												
Amino sugar and nucleotide sugar metabolism												
Caffeine metabolism												
D-Glutamine and D-glutamate metabolism												
Pentose phosphate pathway												
Pyruvate metabolism												
Valine, leucine and isoleucine degradation												
Vitamin B6 metabolism												
Galactose metabolism												
Glycerolipid metabolism												
Glycerophospholipid metabolism												
Sphingolipid metabolism												
Sulfur metabolism												

Figure 2.2: Pathways enrichment analysis based on metabolite concentration differences in urine of ASD patients from 12 different studies.

included amino acids, such as glycine, serine, threonine, alanine, β -alanine, histidine, glutamyl amino acids, and taurine, which were decreased in ASD compared to control group. The level of antioxidant carnosine was also slightly reduced in ASD. Additionally, the authors demonstrated that metabolites from gut bacteria metabolism of amino acids, carbohydrates, and bile acids, were also altered in ASD group. In particular, concentration levels of 2-(4-hydroxyphenyl)propionate and taurocholate sulfate were increased, while the levels of 3-(3-hydroxyphenyl)propionate and 5-aminopentanoic acid were significantly lower in ASD compared to controls.

Changes in purine metabolism pathway, pyrimidine metabolism pathway, and changes in several pathways involved in the metabolism of tyrosine, asparagine, tryptophan, and arginine were further identified in the urine of 30 ASD patients, and 30 age-matched healthy controls using the combination of NMR and LC-MS approaches [Diémé et al., 2015]. Data obtained using six analytical modalities (two NMR and four LC-MS techniques) was divided into a training set consisting of 46 samples, which was used to build OPLS-DA model to find significant features that separate ASD patients from healthy controls and validation set consisting of 16 samples, which was used for evaluation of predictions of OPLS-DA models. The authors identified 19 metabolites that significantly contributed towards separation of ASD and control samples, including methyl guanidine, dihydrouracil, and desaminotyrosine decreased in ASD, and indoxyl sulfate, N-acetyl-L-arginine, glucuronic acid, and phenylacetylglutamine increased in ASD compared to controls.

Most recently, a study employing a combination of NMR and LC-MS approaches to investigate metabolites in urine of 40 ASD patients, and 40 age-matched healthy individuals reported concentration differences for metabolites related to oxidative stress, including glutathione metabolism, changes in cysteine, methionine, arginine and proline metabolism pathways, as well as changes in carbohydrate metabolism, including metabolism of propanoate, citrate, and sugars [Bitar et al., 2018]. They divided the combined NMR/LC-MS dataset into a training set (50 samples) used to build the OPLS-DA model to identify significant features, and a validation set (30 samples) used to evaluate the performance of the classifier. Based on OPLS-DA model, the authors identified 27 highly contributing metabolites that separated ASD patients from healthy controls. The detected differences included alteration of amino acid metabolism, in particular, glycine, serine and threonine metabolism, phenylalanine metabolism, glutamate, arginine and proline metabolism, histidine metabolism, cysteine, and methionine metabolism; carbohydrate metabolism, including 2-hydroxybutyric acid and citric acid, which were decreased in ASD patients compared to controls, and purine metabolism, in particular, 5-aminoimidazole-4-carboxamide, an intermediate metabolite in

purine biosynthesis increased in ASD, and guanine decreased in ASD. Additionally, the authors showed that the concentration of trigonelline, which is formed by methylation of the nitrogen atom of niacin (vitamin B₃), was increased in ASD.

Based on the published autism studies, we can conclude that the highest number of altered metabolic pathways was detected in the studies in which the authors used large numbers of samples and combined results from multiple analytical platforms (LC-MS, GC-MS, NMR). This might be due to the fact that autism patients represent a rather heterogeneous group, so if the larger number of samples is used in the study more changes, we will be able to pass the significance cutoff due to the statistical power increase. That's why to get a more systematic picture of ASD-related changes we may need to increase the number of samples and combine results from multiple analytical platforms.

Based on pathway enrichment analysis that was performed for each study separately, we were able to identify 18 pathways, which were enriched in more than a half of autism metabolic studies. Those pathways can be divided into the following groups: amino acid metabolism (phenylalanine metabolism; glycine, serine and threonine metabolism; tryptophan metabolism; alanine, aspartate and glutamate metabolism; histidine metabolism; lysine degradation; tyrosine metabolism), sugar catabolism pathways (propanoate metabolism; TCA cycle) and also glutathione metabolism and purine metabolism. Those pathways are particularly interesting, as they can be used as targets for drug design to alleviate the ASD-related condition. For example, glutathione concentration is reduced in autism patients, now glutathione precursor might be prescribed to autism patients as a nutritional supplement.

The first study investigating metabolite concentration differences in blood identified changes associated with mitochondrial dysfunction, as well as various metabolic pathway changes, such as disruption of TCA cycle in plasma samples of 52 children diagnosed with ASD and 30 age-matched healthy children using five mass spectrometry-based methods [West et al., 2014]. The authors first divided ASD patients and controls into a training set and validation set. Based on the training set, they trained SVM and PLS classification model to distinguish ASD patients from controls using 179 selected features. Next, with a 21-sample independent validation set, they tested their classifier and obtained the best prediction, with a maximum accuracy of 81% based on 80 feature SVM model. Furthermore, the chemical structures of metabolites were confirmed using tandem mass spectrometry. They reported the following alterations in ASD patients: TCA cycle, including citric acid (decreased in ASD) and succinic acid (increased ASD), and creatinine, the branched-chain amino acid - isoleucine, were reduced in ASD, while glutaric acid, 3-aminoisobutyric acid, dehydroepiandrosterone sulfate, aspartate and glutamate levels in blood were elevated in ASD. Moreover, the authors

demonstrated that homocitrulline was decreased in ASD, which was not previously described elsewhere.

A more recent study identified concentration differences of 17 metabolites in blood plasma of 73 ASD patients. Of them, 11 metabolites, additionally validated in an independent cohort, including sphingosine-1-phosphate and docosahexaenoic acid, were significant predictors of ASD, as they consistently were in top ranks according to multiple logistic regression models [Wang et al., 2016].

The only study conducted in brain identified concentration differences of 37 metabolites in the cerebellum of 11 ASD patients and 11 controls using LC-MS. These differences were not enriched in any biological pathway [Graham et al., 2016]. The highest magnitude of metabolic differences with higher concentration in ASD patients compared to controls was observed for 5,6-dihydrouridine, asparaginyln-hydroxyproline, phosphatidylglycerophosphate(16:1: ω 7/18:0) and CDP-diacylglycerol(18:1/18:2). By contrast, 2-octanoyl carnitine, phytyl diphosphate, and N-lignoceroylsphingosine demonstrated lower concentrations in ASD patients compared to controls.

In addition to studies assessing the concentration levels of multiple metabolites, others focused on particular compounds, such as markers of mitochondrial dysfunction that are believed to be associated with disease [Eapen, 2011, Rossignol and Frye, 2012, El-Ansary et al., 2017]. These studies reported an elevated concentration of glutamate [Shinohe et al., 2006, Shimmura et al., 2011], as well as glycolysis products, such as lactate and pyruvate, in the serum of ASD patients [Laszlo et al., 1994]. Other findings included a decreased concentration of carnitine, the fatty acid carrier from the cytosol to the mitochondria [Filipek et al., 2004], and glutathione, the reported critical reactive oxygen species neutralizer, in ASD patients' blood [James et al., 2004]. By contrast, the concentration of palmitate, one of the principal energy sources for mitochondria, was increased in ASD plasma samples [Pastural et al., 2009].

2.5 Lipids of the cell and their involvement in neurological disorder

Lipids are essential components of cell plasma membranes. They also function as energy storage blocks, signaling molecules, and protein recruitment platforms in all cells and tissues, especially in the brain [Simons and Toomre, 2000, Harayama and Riezman, 2018]. Their critical role in cell functioning in the brain and other tissues is further supported by their association with neurological and metabolic disorders, such as ASD, schizophrenia, Alzheimer

disease, Parkinson disease, as well as diabetes and cancer [Kim et al., 2010, El-Ansary et al., 2011, Kaddurah-Daouk et al., 2007, Orešič et al., 2012, Solberg et al., 2016, Han et al., 2002, Wenk, 2005, Adibhatla et al., 2006, Colsch et al., 2008, Ariga et al., 2008, Haughey et al., 2010, Lamari et al., 2013]. Here, I will focus on the known lipid concentration changes in ASD and schizophrenia. Similar to metabolites, most of the lipidomics studies were performed using urine and blood plasma samples.

2.5.1 Diversity of membrane lipid chemical structures

Functions of lipids rely on their chemical structures. LIPID MAPS consortium has classified lipids into eight classes: fatty acyls, glycerolipids, glycerophospholipids, sphingolipids, sterol lipids, prenol lipids, saccharolipids, and polyketides [Fahy et al., 2009].

The first group is the fatty acyls, which constitute a group of lipid molecules synthesized by elongation of an acetyl-CoA with malonyl-CoA (or methylmalonyl-CoA). Fatty acyls are characterized by a repeating series of methylene groups. Fatty acyls are the basic building blocks of more complex lipids such as glycerophospholipids, glycerolipids, sphingolipids, and glycolipids [Fahy et al., 2009].

The diversity of glycerophospholipids arises from the combination of two fatty acids, the linkage, and the head group (Figure 2.3 [Harayama and Riezman, 2018]). The fatty acids are exceedingly diverse: they can differ in length of the carbon chain and can have multiple double bonds (polyunsaturated), one double bond (monounsaturated), and no double bonds (saturated). The *sn-1* fatty acid tends to be saturated or monounsaturated, while the *sn-2* fatty acid is more often monounsaturated or polyunsaturated [Yamashita et al., 2014]. The nomenclature of free fatty acids is CN:DB, where CN shows the number of carbons in the chain and DB corresponds to the number of double bonds. Additionally, the position of the first double bond in the fatty acid chain is sometimes added, for example, omega-3 (ω 3) and omega-6 (ω 6), in which double bond is at 3rd or 6th carbon atom in the tail of the chain. The head group consists of phosphate and alcohol. The alcohol group defines the lipid class name, for example, choline (phosphatidylcholine, PC), ethanolamine (phosphatidylethanolamine, PE), serine (phosphatidylserine, PS), inositol (phosphatidylinositol, PI) and glycerol (phosphatidylglycerol, PG). Two examples of glycerophospholipids with less diversity are PS tending to have stearic acid (C18:0) at the *sn-1* position, and PI tending to have stearyl (C18:0) and arachidonoyl-acyl (C20:4: ω 6) chains [Hicks et al., 2006, Yamashita et al., 2014].

Another class of lipids is glycerolipids. Glycerolipids are similar to glycerophospholipids, except for the phosphatide group, which is substituted by a hydroxyl group. The diversity

of glycerolipids is due to the number, length, and type of fatty acid chains. If glycerolipids have three fatty acid chains, they are named triglycerides (TG), two - diglycerides (DG), one - monoglycerides (MG).

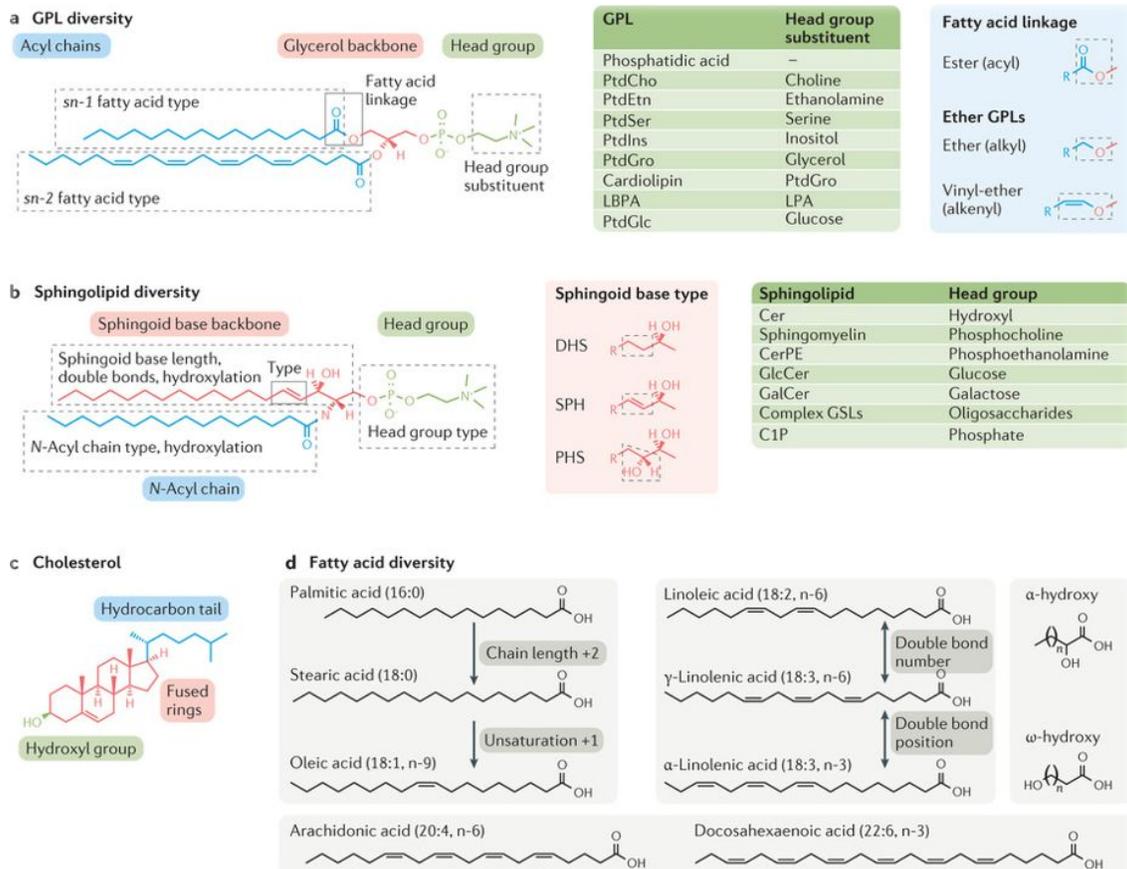


Figure 2.3: Diversity of lipids composing the membrane: glycerophospholipids, sphingolipids, cholesterol and free fatty acids (reproduced with permission from Nature Reviews Molecular Cell Biology [Harayama and Riezman, 2018]).

The diversity of sphingolipids originates from the length and type of the sphingolipid base, N-acyl chain, and the head group (Figure 2.3 [Harayama and Riezman, 2018]). N-acyl chains of sphingolipids are the same *sn-1* and *sn-2* fatty acids that form glycerophospholipids but N-acyl chains tend to be more saturated and can be longer than in *sn-1* and *sn-2* fatty acids [Grösch et al., 2012]. The head contains one of the following groups: hydroxyl (ceramides), phosphocholine (sphingomyelin), phosphoethanolamine (ceramide phosphoethanolamines). Besides that, the head group can include various oligosaccharides (glycosphingolipids).

Another class of lipids is sterols, which contains a common steroid core of a fused four-ring structure with a hydrocarbon side chain and an alcohol group. This group can be subdivided

into cholesterol and derivatives, steroids, secosteroids, bile acids, and their derivatives. The primary sterol lipid in mammals is cholesterol that is an essential component in cellular membranes (Figure 2.3 [Harayama and Riezman, 2018]).

The remaining three classes: prenols, saccharolipids, and polyketides, are less studied. The prenol lipids are synthesized from the five-carbon precursors isopentenyl diphosphate and dimethylallyl diphosphate that are produced mainly via the mevalonic acid pathway [Han, 2016]. To the saccharolipids class belongs the lipids in which fatty acids are linked directly to a sugar backbone (a glycerol backbone of glycerolipids or glycerophospholipids is substituted by a sugar) [Han, 2016]. The polyketides are a class of lipids from a plant and microbial sources, which are synthesized by polymerization of acetyl and propionyl groups [Han, 2016].

2.5.2 Lipid functionality

Lipids can modulate their functions through changes in membrane lipid composition, which can lead to changes in membrane physicochemical properties (Figure 2.4) [Harayama and Riezman, 2018].

All lipid characteristics, including fatty acid chains and the head group, affect the shape of the lipid bilayer and the spontaneous curvature of the membrane [Ernst et al., 2016]. Membrane remodeling processes include fusion and fission. In the process of fusion, the membrane has negative spontaneous curvature, which is obtained by introducing additional cone-shaped lipids (small head group compared to the large fatty acid group), such as PE. Increasing the number of inverted cone-shaped lipids, such as PI, giving membrane a positive spontaneous curvature property, controls the process of fission [Dowhan and Bogdanov, 2002].

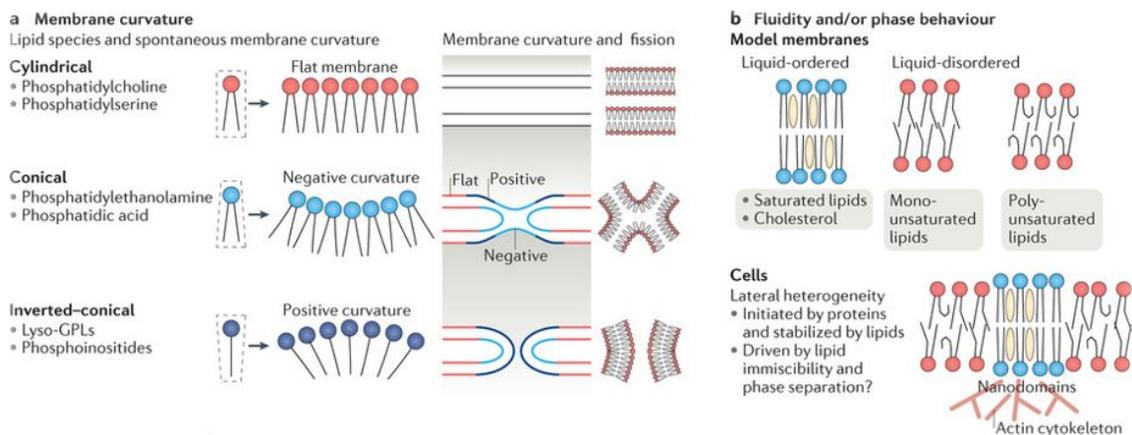


Figure 2.4: Membrane properties affected by lipid composition (reproduced with permission from Nature Reviews Molecular Cell Biology [Harayama and Riezman, 2018]).

Lipids with long SFA make membranes thicker and less fluid, while MUFA and PUFA make membranes more fluid because double bonds make bends in fatty acid chains and repulse lipids from each other [van Meer et al., 2008]. Therefore, membrane fluidity can be controlled by the ratio of saturated/unsaturated lipids. Another critical aspect affecting membrane fluidity is lipid-lipid and lipid-protein interaction.

In most cases, lipid functionality is mediated not by direct changes of membrane properties, but rather through the proteins with which lipids interact (Figure 2.5) [Harayama and Riezman, 2018].

Many lipids drag proteins containing lipid-binding domains towards the membrane. One of the well-studied examples of lipids that interact with lipid-binding proteins is PI. They have an inositol group, which can regulate its binding activity through phosphorylation/dephosphorylation process [De Craene et al., 2017]. Besides that, it was shown that the specificity of binding could be achieved by cooperative binding with various lipids [Vonkova et al., 2015].

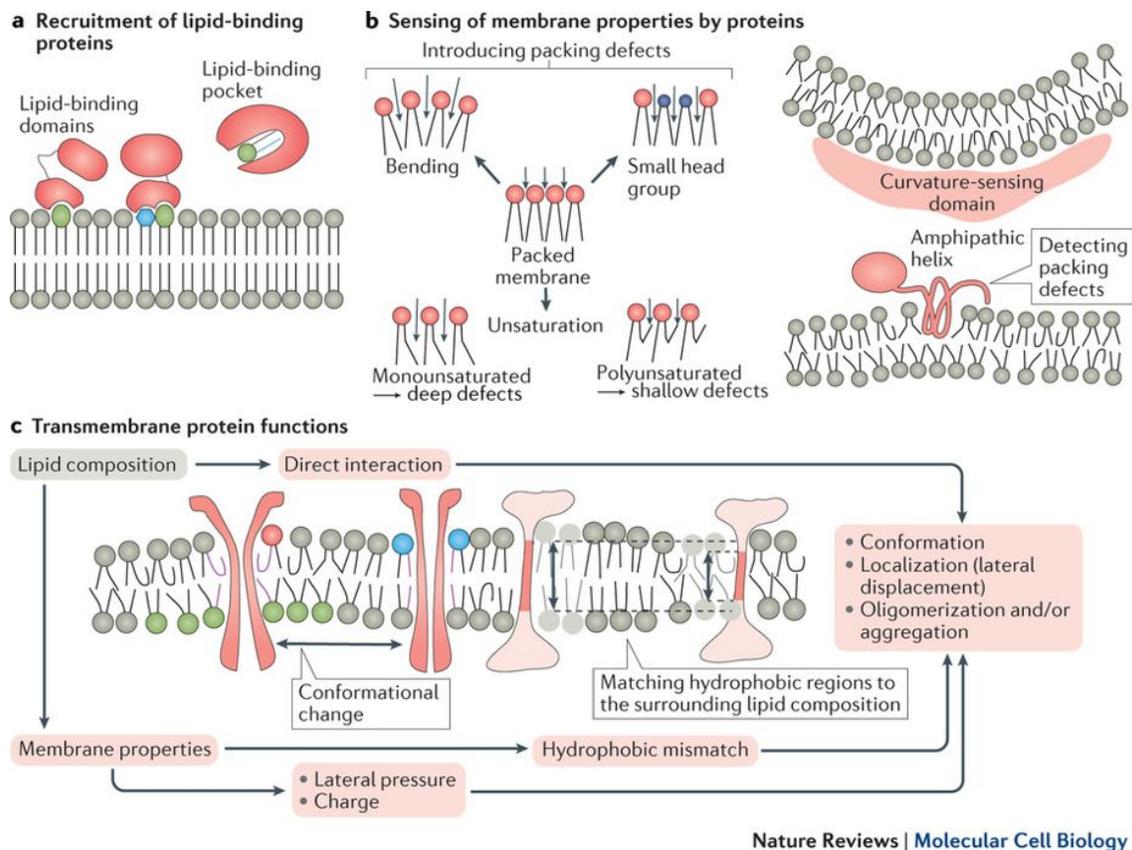


Figure 2.5: Lipids mediate their function through lipid-protein interactions (reproduced with permission from Nature Reviews Molecular Cell Biology [Harayama and Riezman, 2018]).

Another type of lipid-protein interaction involves proteins that have lipid-binding pockets. Some nuclear receptors can regulate their activity by binding membrane lipids, which alter protein structure. PC, PI, and sphingolipids are among lipid classes that participate in such interactions [Crowder et al., 2017]. Lipid transfer proteins responsible for spatiotemporal distribution of membrane lipids via non-vesicular transfer between biological membranes also contain lipid-binding pockets. Besides transfer function, lipid transfer proteins can act as chaperones and present lipids to other proteins, for example, nuclear receptors. They can also serve as lipid sensors that perceive fluctuations of levels of specific lipids [Chiapparino et al., 2016].

Another class of protein-lipid interaction includes proteins which do not bind directly to individual lipids, but sense membrane properties. For instance, the proteins that can sense the curvature contain bin-amphiphysin-rvs (BAR) domain. BAR domains consist of dimeric alpha-helices that bind to highly curved negatively charged membranes and stabilize membrane curvature [Peter, 2004]. Another important membrane property is a packing defect, which is characterized by the hydrophobic fatty acid part of the lipid bilayer being exposed to the aqueous environment. Lipid composition can cause such defects. For example, high levels of PUFA lead to greater packing defects compared to SFA. Proteins with amphipathic helices, such as the ALPS motif and the N-terminal region of α -synuclein, bind to packing defects and stabilize membrane curvature [Antonny, 2011].

The lipid composition of the membrane can affect functions of transmembrane proteins by changing its conformation by direct lipid-protein binding or indirectly through membrane properties. Membrane thickness can also affect lateral protein distribution because proteins need to adjust the length of transmembrane domains according to local lipid composition [Harayama and Riezman, 2018].

2.5.3 Lipidome alterations in autism

Multiple studies investigating alterations of lipid metabolism in ASD patients focused on PUFA, as they compose 20% of the dry weight of the brain, and particularly on ω 3 docosahexaenoic acid (C22:6: ω 3; DHA) and ω 6 arachidonic acid (C20:4: ω 6; AA), the two most abundant PUFA in the brain. Multiple studies focused on the investigation of PUFA concentration levels in blood plasma and red blood cells of ASD patients [Vancassel et al., 2001, Bell et al., 2004, Bu et al., 2006, Wiest et al., 2009, Gordon Bell et al., 2009], as it was shown that the levels of fatty acids, particularly DHA, in both blood plasma and red blood cells correlate well with lipid levels in brain [Kuratko and Salem, 2009].

In one study, representing the most extensive investigation to date, authors measured eight different lipid classes CE, FC, PC, PE, LyPC, TG, DG, and FA in blood plasma of ASD ($n = 153$) and typically developing ($n = 97$) children [Wiest et al., 2009]. They showed that, among cholesterol esters, the two essential fatty acids, C18:3: ω 3 and C18:2: ω 6 were decreased in ASD compared to control. For free cholesterol, there were no significant changes between ASD and control. In phosphatidylcholine class, only C20:2: ω 6 was increased in the ASD compared to control group. In phosphatidylethanolamine class C20:4: ω 3, C20:2: ω 6, C20:3: ω 9, C18:1: ω 9, and C18:1: ω 7 were increased in ASD, in contrast to C22:6: ω 3 (DHA), and C24:1: ω 9, which were decreased in ASD compared to control. In lysophosphatidylcholine class, the concentrations of C20:4: ω 3 and C18:1: ω 9 were higher in ASD compared to control. In triglycerides, C22:2: ω 6 was increased while C15:0 and C22:5: ω 6 were decreased in ASD compared to control. In diglycerides, C22:1: ω 9 and C18:0 were lower and C14:0 and C20:3: ω 6 were higher in ASD compared to control. In free fatty acids class, the concentrations of C14:1: ω 5 (myristoleic acid) and C20:4: ω 6 (AA) were decreased in ASD.

A more detailed analysis of free fatty acid concentrations in blood plasma was based on the comparison of 26 ASD patients and 26 age-matched controls between 4 and 12 years of age [El-Ansary et al., 2011]. The authors demonstrate a decrease in the concentration of most PUFA, which agrees with other studies, and an increase of saturated fatty acids, except for propionic acid, in ASD patients compared to controls.

Another study investigated triglycerides and cholesterol concentration levels in blood plasma of 29 boys diagnosed with ASD and 29 unaffected boys from 7 to 12 years of age [Kim et al., 2010]. The authors showed that average TG concentration levels were increased in ASD compared to control whereas the average HDL-C level was decreased in ASD. They also demonstrated that there is no difference in concentration levels of TC and LDL-C between ASD and control groups.

In the largest study conducted to date, the authors compared red blood cells PUFA levels in ASD and control individuals. In this study, ASD individuals ($n = 121$) and typically developing controls ($n = 110$) with ages between 3 and 17 years were investigated [Brigandi et al., 2015]. The authors demonstrated that children with ASD had lower DHA and AA levels than controls in red blood cells. Additionally, they showed that accumulative levels of all measured ω 6 and ω 3 fatty acids were lower in ASD individuals compared to controls. Moreover, authors measured Prostaglandin E2 (PGE2) concentration, which is a metabolite of AA, in 20 ASD individuals and 20 matching controls and demonstrated that PGE2 was increased in ASD compared to controls [Brigandi et al., 2015].

Most recently, AA, DHA and Eicosapentaenoic acid (C20:5: ω 3; EPA) were measured in red blood cells in 85 ASD individuals and 79 controls with ages between 3 and 17 years. The authors demonstrated that children diagnosed with ASD had lower DHA, EPA and AA levels, higher AA/EPA ratio, and lower ω 3/ ω 6 ratio than controls [Parletta et al., 2016]. All these observations agree with the results of another study [Brigandi et al., 2015].

2.5.4 Lipidome alterations in schizophrenia

A pioneering study mapped global lipid changes in schizophrenia before and after treatment with antipsychotic drugs (50 patients, 18-60 years of age) and 16 matching controls [Kaddurah-Daouk et al., 2007] in blood plasma. The authors were able to quantify over 300 lipids from seven lipid classes: CE, DG, FA, LyPC, PC, PE and TG. None of these lipid classes differed significantly between schizophrenia patients before treatment and controls, except for PE class, where MUFA and ω 6 PUFA were decreased in schizophrenia patients compared to controls. The authors demonstrated that PC, PE, and TG were increased while FA were decreased in olanzapine-treated schizophrenia patients compared to non-treated patients. Usage of the second drug, risperidone, demonstrated rather similar results: increased PE, PC, and LyPC concentration levels and decreased FA. The third drug, aripiprazole, demonstrated an increase in PE lipid class treated compared to non-treated schizophrenia patients. The common effect of drugs is an increase of PE lipid class, as by taking these drugs patients elevate their PE level, which is lower in schizophrenia patients compared to controls. This type of experiment is particularly interesting, as using this approach an optimal drug combination could potentially be found for a particular patient, i.e. drugs that change only disease altered lipid classes and do not have off-target effects.

Another study that measured lipids in blood plasma was based on 19 patients with schizophrenia and 34 age and gender-matched healthy twins as controls [Orešič et al., 2012]. Using LC-MS platform, the authors were able to identify 250 lipids belonging to PC, PE, TG, FA, LyPC, and SM. They demonstrated that LyPC were decreased in schizophrenia patients compared to controls, while TG levels were elevated.

A more detailed analysis of lipid composition of the red blood cells was performed in 74 schizophrenia patients and 40 control individuals [Tessier et al., 2016]. The authors measured all major lipid classes: PC, PS, SM, PE and PE plasmalogen using LC-MS/MS approach. They demonstrated that SM concentration levels have a clear bimodal distribution in schizophrenia patients (41/33), although a small number of controls showed a low concentration level of SM as well (7/33). The first group of schizophrenia patients characterized by low SM membrane content demonstrated a decrease in PE, PC, PE plasmalogen while levels

of PS were increased compared to controls. The second group of schizophrenia patients was similar to controls regarding membrane SM composition but levels of PE and plasmalogen PE were decreased while PC and PS were increased compared to controls. Additionally, authors showed that the group of schizophrenia patients with decreased SM levels was characterized by significantly more severe psychopathology and impaired cognitive performance.

Most recently, free fatty acids were measured in blood serum of 110 schizophrenia patients and 109 controls, using UPLC-QTOF-MS platform [Yang et al., 2017]. The authors detected 47 free fatty acids that were divided into 8 groups: SFA, MUFA, ω 6 PUFA, ω 3 PUFA, TFA, BCFA, OCFA and others. All MUFA and ω 6 PUFA were increased in schizophrenic patients compared to controls, including the following MUFA: C14:1: ω 5, C16:1: ω 7, C18:1: ω 9, C20:1: ω 9, C22:1: ω 9, C24:1: ω 9 and ω 6 PUFA: C20:2: ω 6, C20:3: ω 6, C20:4: ω 6, C22:2: ω 6, C22:4: ω 6. As for the remaining classes, even though the majority of detected lipids (including C22:5: ω 3, C17:1: ω 7, C19:1: ω 9, C16:0) demonstrated elevated concentration levels in schizophrenia patients compared to controls, there were some lipids that showed decrease in schizophrenia patients compared to controls, for example, C24:0.

The pioneering study investigating lipidome composition of white and gray matter in the prefrontal cortex (Brodmann area 9) was based on 15 schizophrenia patients and 15 healthy controls [Schwarz et al., 2008]. The authors found alterations in levels of FA, PC and ceramides in gray and white matter of schizophrenia patients compared to controls. Specifically, PC were increased in gray matter but decreased in white matter, FAs were increased in both gray and white matter, while ceramides were increased only in white matter. Additionally, to compare the results obtained in the prefrontal cortex with lipid composition of blood, authors measured red blood cells in 20 schizophrenia patients and 20 healthy controls. They demonstrated that alterations in concentrations of free fatty acids, as well as ceramide, were in agreement with the differences found in the cortex.

2.6 Evolutionary models at different molecular levels

2.6.1 Evolutionary studies at the level of genome and transcriptome

Genome sequence and phenotype differences among species are relatively well studied. This knowledge forms the foundation for evolutionary models that explain observed genetic and phenotypic divergence and help to identify genetic regions or phenotypic traits influenced by a specific type of evolutionary selection. Under most, though not all, current models, the majority of genetic differences observed among species are considered to be evolutionarily neutral, that is, have no effect on the individual's survival and reproductive abilities, while

a few differences are adaptive, that is, increase the individual's fitness [Otto, 2000, Eyre-Walker, 2006, Hahn, 2008, Orr, 2009]. Genetic changes with substantial deleterious effects are removed by purifying selection and are not observed in comparisons among species. Accordingly, some of the main features of genetic evolution include: 1) an overwhelming correlation between phylogenetic distances and genetic distances; 2) reduced diversity (variation within species) and divergence (differences between species) of functional sequences compared with nonfunctional ones; and 3) reduced diversity but increased divergence of sequences containing adaptive variants.

In contrast to genetic differences, most changes that occur at the level of the phenotype affect individuals' survival and reproductive chances and only a few changes are neutral [Nei, 2007, Steiner and Tuljapurkar, 2012]. While the genetic information largely determines the phenotype, several intermediate steps enable the transition of this information to the phenotype level. These steps include changes in epigenetic modification levels of DNA and histone proteins, RNA and protein expression levels, as well as concentrations of small molecules, metabolites, and lipids, that participate in all biological processes. Some of these steps, including RNA expression levels, have been examined, resulting in evolutionary models parallel to the one based on genetic data [Wray et al., 2003, Jordan et al., 2005, Stern and Orgogozo, 2008].

One of the large-scale studies investigating the evolution of gene expression levels was done in 6 tissues (liver, heart, kidney, testis, brain, cerebellum) of 9 species representing placental mammals, marsupials, monotremes and one bird species (red jungle fowl), which served as the evolutionary outgroup [Brawand et al., 2011]. To estimate rates of expression change in different evolutionary lineages in different organs, expression-based phylogenetic trees were built. The total branch lengths of these trees varied among tissues, with two neural tissues having the shortest branches and testis having the longest ones. Based on these observations, the authors suggested that transcriptome change in nervous tissues might be largely influenced by purifying selection, in contrast to testes, which showed signs of positive selection. Next, pairwise comparisons between all species were performed and confirmed an overall increase in gene expression divergence with evolutionary time. Additionally, the authors estimated gene expression changes on different chromosomes and showed that gene expression evolution was faster on X chromosome compared to autosomes in the common ancestor of therian mammals, in contrast to eutherian branches, where the evolution rate on X chromosome was similar to autosomes. Additionally, using a modified OU model applied to each gene in each tissue separately, the authors identified genes with potentially adaptive expression changes.

Most recently, gene expression was measured in 3 tissues (liver, kidney, and brain) of 33 diverse species of mammals [Fushan et al., 2015]. The primary focus of this work was the estimation of the relationship between gene expression levels of different species and their life histories, such as time to maturity, oxygen consumption and maximum lifespan, in particular. The authors implemented non-phylogenetic OLS method to investigate the relationship between gene expression and organism trait, as it was shown that life history variation departs from the model of neutral evolution. The analyses provided evidence that life histories can explain up to 20% of gene expression variation between species. Next, Pagel's λ model [Pagel, 1999] was applied to gene expression levels for each gene separately to estimate the proportion of genes with expression variation between species explained by drift. The highest percentage of genes that followed phylogeny, based on Pagel's λ model, was observed in the brain. To estimate the number of genes with expression variation explained by stabilizing constraints, the authors modeled each gene as an OU process with stepwise Akaike Information Criterion [Ingram and Mahler, 2013]. The highest proportion of genes having expression consistent with the OU model was also observed in the brain.

2.6.2 Evolutionary studies at the level of metabolome and lipidome

While the evolutionary models describing the changes at the genomic and transcriptomics levels are relatively well established, the number of studies that investigate the evolutionary changes at yet another intermediate step connecting the genome with the phenotype – the level of metabolite and lipid concentration changes - is much smaller.

In a pioneering study, metabolite concentrations were measured in the prefrontal cortex of humans (schizophrenia patients and controls), chimpanzees and macaques using proton-NMR spectroscopy. There was an overlap between metabolite changes in schizophrenia and evolutionary changes observed on the human lineage [Khaitovich et al., 2008]. The authors compared metabolite concentrations between schizophrenia patients and control individuals using t-test and identified nine metabolites, including creatine, lactate, glycine, acetate, choline, phosphocholine, glycerophosphocholine to be altered between those two groups. Next, they measured changes in concentration on the human and chimpanzee lineages in the two metabolite groups (changed and non-changed in schizophrenia patients compared to controls) using the rhesus macaque as an outgroup. They showed that the ratio of the human to chimpanzee branch length is more than three times higher for the changed in schizophrenia metabolite group than for the non-changed metabolite group. This work has demonstrated that metabolic changes are essential for establishment and maintenance of the human-specific cognitive abilities.

Another study assessed the metabolome composition in the prefrontal and cerebellar cortex of humans, chimpanzees, and macaques during postnatal development using GC-MS [Fu et al., 2011]. First, choosing a suitable polynomial model for age and then applying ANCOVA approach the authors were able to identify metabolites with concentration differences between species also demonstrating changes with age. They showed that there was a four-fold excess of human-specific metabolic changes in the prefrontal cortex compared to cerebellar cortex. Specifically, the human-specific metabolic changes in the prefrontal cortex were enriched in the following metabolic pathways: long-term potentiation, neuroactive ligand-receptor interaction, alanine, aspartate, and glutamate metabolism, and β -alanine metabolism.

Most recently, a study employing a combination of GC-MS and LC-MS approaches assessed the concentration of metabolites in muscle, kidney, prefrontal cortex, cerebellar cortex and primary visual cortex in humans, chimpanzee, macaques and mouse [Bozek et al., 2014]. They performed ANOVA to find metabolites with significant concentration differences among the four species within each of the five tissues, and within the combined brain tissues. They demonstrated that there is an excess of human-specific metabolic divergence in the skeletal muscle, prefrontal cortex, and in the brain as a whole, but not in the primary visual cortex or the kidney. Human-specific metabolic changes in the prefrontal cortex were associated with translational metabolism and neurotransmitter signaling, while human-specific changes in muscle showed enrichment in carbohydrate metabolism energy and amino acid metabolism. Similar to the previous studies, the general evolution of metabolite levels was not studied because the primary focus of the work was to investigate human-specific concentration differences in the brain, particularly in the prefrontal cortex and muscle.

The only study where general trends of metabolite and lipid concentration evolution were studied, which is the most complete metabolome study up to date, detected concentration levels of metabolites in 4 tissues (brain, heart, kidney, liver) from 26 mammals using GC-MS data [Ma et al., 2015]. The authors tested if metabolite concentration levels reflected the phylogenetic tree based on proportion of metabolites with high phylogenetic signals, using Pagel's λ [Pagel, 1999] and Blomberg's K [Blomberg et al., 2003] in each tissue separately. Additionally, average tip-to-root branch length for each tissue was calculated to estimate the metabolite divergence. Based on these statistics, the authors showed that brain had the largest proportion of metabolites with high phylogenetic signals and the shortest tip-to-root branch lengths compared to the remaining three tissues, arguing that the brain evolved according to phylogenetic tree, in contrast to other tissues that did not follow phylogeny as their metabolite concentration levels diverged more between species.

Similarly, there are only a few studies investigating lipid concentration level differences at the evolutionary level [Bozek et al., 2015, Li et al., 2017, Bozek et al., 2017].

The pioneering study that estimated the lipidome diversity in muscle, kidney, prefrontal cortex, cerebellar cortex and primary visual cortex of humans, chimpanzees, macaques, and mice, was using LC-MS approach [Bozek et al., 2015]. The authors used ANOVA to identify lipids with concentration levels changing significantly among species or tissues. First, it was shown that in humans the majority of tissue differences is specific to the brain, and all lipidome differences were divided into brain-enriched (high concentration levels in the brain compared to other tissues) and brain-depleted (low concentration levels in the brain compared to other tissues) lipids. Next, the authors demonstrated that lipid concentrations measured in chimpanzees, macaques, and mice resembled the lipidome profiles of human tissues. Finally, the ratios of human-specific to chimpanzee-specific lipids for brain-enriched and brain-depleted groups were calculated. Based on these ratios, it was concluded that brain-enriched lipids evolved approximately four times faster than brain-depleted lipids of non-neural tissues. Even though the authors tried to address evolutionary questions in this study [Bozek et al., 2015], the species set was limited to humans, chimpanzees, rhesus macaques, and mice, and therefore did not provide sufficient data for the general analysis of lipid concentration evolution. Thus, in this study, the authors focused on lipid concentration changes specific to the human brain.

An even smaller set of species was used in the following study [Li et al., 2017] - humans, chimpanzees, and macaques. This study focused on developmental changes of the lipidome composition of the prefrontal cortex using LC-MS. The authors used ANCOVA approach to demonstrate that the majority of lipids change their concentration significantly along the lifespan; specifically, more than 60% of these changes occur before adulthood. In addition to these observations, the authors demonstrate that there is a two-fold excess of human-specific lipids compared to the number of chimpanzee-specific lipids and the greatest amount of both human- and chimpanzee-specific differences occurs in early development, while the greatest divergence of human- and chimpanzee-specific lipids is observed at the human age of 20 to 35. Similar to the previous study, the general evolution of lipid levels was not studied as the primary focus of the work was on human-specific age-related concentration differences in the prefrontal cortex.

The most extensive study to date compared lipid concentration levels in six tissues (liver, heart, muscle, kidney, cortex, cerebellum) from 35 species representing three mammalian clades: primates, rodents, and bats using LC-MS [Bozek et al., 2017]. Using logistic regression with the elastic net penalty, the authors identified lipids that might predict the

maximum lifespan of species in 6 tissues. Interestingly, these lipids overlapped significantly between brain cortex and cerebellum, as well as among three non-neural tissues: liver, muscle, and kidney. Next, the authors measured dN/dS ratios for enzymes linked to lipid predictors of maximal lifespan and the remaining lipids in short-living and long-living species. They demonstrated that enzymes linked to lipid predictors of lifespan had a significantly lower amino acid substitution rate in contrast to enzymes linked to the remaining lipids in long-living species but not in short-living ones, providing evidence of increased evolutionary pressure in metabolic processes of long-living species. Therefore, the general evolution of lipids was not studied in [Bozek et al., 2017], except for the evolution of one small and specific group of lipids - lipid predictors of maximal lifespan.

Chapter 3

Metabolome signature of autism in the human prefrontal cortex

In this chapter, we investigated concentration levels of 1,366 metabolites in the prefrontal cortex grey matter from 40 healthy humans and 32 autism patients, as well as 40 chimpanzees and 40 macaques, with ages spanning the entire postnatal cortical development. We identified numerous metabolic changes distinguishing autism patients from the controls. Many of these differences overlapped with the pathways previously detected in urine and blood. Furthermore, several of these pathways contained an excess of metabolic changes unique to humans, indicating disruption of the evolutionary novel cortical mechanisms in ASD.

3.1 Materials and Methods

3.1.1 Samples

This study was reviewed and approved by the Institutional Animal Care and Use Ethics Committee at the Shanghai Institute for Biological Sciences, CAS. Informed consent for the use of human tissues for research was obtained in writing from all donors or their next of kin. All non-human primates used in this study suffered sudden deaths for reasons other than their participation in this study and without any relation to the tissue used.

We used prefrontal cortex (PFC) samples dissected from the frozen postmortem brains of 40 cognitively unaffected human controls (0-62 years old), 32 ASD cases (2-60 years old), 40 chimpanzees (0-43 years old), and 40 rhesus macaques (0-21 years old). Special care was taken to dissect gray matter only.

Control human samples were obtained from the NICHD Brain and Tissue Bank for Developmental Disorders at the University of Maryland, USA, the Maryland Brain Collection Center, Maryland, USA and the Harvard Brain Tissue Resource Center. ASD samples were obtained from the NICHD Brain and Tissue Bank for Developmental Disorders and the Harvard Brain Tissue Resource Center. All the control and ASD brain samples used in this study were also part of recently published transcriptomic [Liu et al., 2016] and lipidomic [Yu et al., 2018] studies.

Chimpanzee samples were obtained from the National Chimpanzee Brain Resource (NS092988), the Alamogordo Primate Facility, New Mexico, USA, the Anthropological Institute and Museum of the University of Zurich-Irchel, Switzerland, the Biomedical Primate Research Centre, the Netherlands, Department of Anthropology, The George Washington University, Washington, DC and Burgers' Zoo in Arnhem, the Netherlands. Rhesus monkey samples were obtained from the Suzhou Experimental Animal Center, China. PFC dissections were made from the frontal part of the superior frontal gyrus. For all samples we preferentially dissected and analyzed gray matter material.

To identify metabolites affected by postmortem delay in primates, we additionally collected two rhesus macaque samples, dissected 5-6 hours after death. Other macaque samples used in this study had PMDs lower than 20 minutes.

3.1.2 Sample preparation

Metabolites were extracted from frozen tissue powder using a methanol:methyl-tert-butyl-ether (1/3 (vol/vol)) solution as described in [Giavalisco et al., 2011]. In brief, 10-15 mg tissue samples were dissected on dry ice from the frozen tissue without thawing. Dissected samples were weighed and transferred to pre-cooled 2 ml round bottom reinforced Precellys tubes containing 2.8 mm zirconia beads. After the addition of a fixed volume (0.5 ml) of pre-cooled (-20 °C) extraction buffer to each tube, we performed two cycles of homogenization using the following parameters: stirring intensity 5,000 rpm, homogenization temperature 7 °C, cycle time 45 second, and 15 second break. Several blank extraction samples were added to the end of each batch (48 samples). The blank samples consisted of the same 2 ml round bottom reinforced Precellys tubes without any sample material subjected to all extraction procedures. After the addition of a fixed volume (0.5 ml) of extraction buffer, the homogenates were vortexed and incubated for 30 minutes at 4 °C on an orbital shaker followed by a 10 minute ultra-sonication in an ice-cooled sonication bath. For each sample, the homogenate was transferred to a pre-labeled 2 ml Eppendorf tube followed by the addition of 700 μ l of an H₂O: methanol (3:1 (vol/vol)) solution containing 0.7 μ g of corticosterone and

ampicillin. Finally, to separate the organic phases from aqueous phases and to precipitate proteins, after a brief vortexing, the homogenates were centrifuged for 10 minutes at 14,000 g at 4 °C. Subsequently, 300 μ l of the lower aqueous phase containing hydrophilic compounds were transferred to a pre-labeled 1.5 ml Eppendorf tube and the solvent was evaporated using a speed vacuum centrifuge at room temperature. Dry metabolic extracts were stored at -80 °C prior to MS analysis. For quality control, we further constructed a pooled sample composed by the aliquots of metabolite extracts from all samples used in the analysis.

3.1.3 Mass spectrometry analysis

The dried extracts were resuspended in 200 μ l of ice-cold 20% aqueous solution of acetonitrile prior to MS analysis. After a brief rigorous vortexing, the samples were incubated for 30 minutes at 4 °C on an orbital shaker followed by a 10 minutes ultra-sonication in an ice-cooled sonication bath and centrifugation for 10 minutes at 14,000 g at 4 °C. For the MS analysis, 40 μ l of supernatant was transferred to 350 μ l autosampler glass vials (Glastechnik Grafenroda, Germany). A chromatography separation of metabolites prior to MS 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. We used two mobile phases for the chromatographic separation. Buffer A was water containing 10 mM ammonium acetate and 0.2 mM ammonium hydroxide: acetonitrile (95/5 (vol/vol)) solution (pH value 8.0), and buffer B was 100% acetonitrile. The gradient separation was: 0 minutes 0% A, 0.01-10 minutes linear gradient from 0% to 100% A, 10-14 minutes 100% A, 14-17 minutes linear gradient from 100% to 0% A, and 17-25 minutes 0% A. The flow rate was set to 500 μ l/minute. The column temperature was maintained at 40 °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. Spray voltage was set to 4.5 kV in positive mode and to 3 kV in negative mode. The remaining MS settings were the same in both ionization modes: S-lens RF level - 70; heated capillary - 250 °C; aux gas heater temperature - 350 °C; sheath gas flow rate - 45 arbitrary units; aux gas flow rate - 10 arbitrary units; sweep gas flow rate - 4 arbitrary units. Full scan resolutions were set to 70,000 at 200 m/z. The full scan target was 106 with a maximum fill time of 50 ms. The spectra were recorded using full scan mode, covering a mass range from 100 to 1500 m/z. For quality control, pooled samples were injected 4 times prior to each ionization mode run in order to condition the

column and after the completion of both ionization modes. Additionally, a pooled sample was injected after every 48th sample injections to assess the sensitivity and retention time consistency of the system, sample reproducibility, and the compound stability over the time of the MS analysis. Sample randomization was performed twice: before the lipid extraction and before the mass spectrometry measurements. Those factors we considered in the process of randomization: diagnosis, age, and species.

3.1.4 Mass spectrometry data processing

MS peaks obtained in positive and negative modes were aligned across samples as described in [Giavalisco et al., 2011, Bozek et al., 2015]. Metabolite identification was based on the computational matching of m/z values to the Human Metabolome Database (HMDB) [Wishart et al., 2012] and the LIPID MAPS Structure Database (LMSD) [Fahy et al., 2009] with a mass tolerance of 10 ppm, allowing [M+H], [M+NH₄], and [M+Na] modifications as possible adducts in positive ionization mode, and [M-H], [M+Formic acid-H], [M-H₂O-H] and [M+Na-2H] modifications in negative ionization mode [Giavalisco et al., 2011, Bozek et al., 2014].

To eliminate the effects of the measurement order on metabolite concentrations, we computed this effect for each metabolite by fitting a support vector regression model with a gaussian kernel and one set of parameters for all metabolites to the \log_2 transformed, centered to the mean = 0, and scaled to standard deviation = 1 concentration values. The resulting functions were clustered using a k-means algorithm. After a visual evaluation, the clusters obviously affected by the measurement order were discarded. The concentration values of the remaining metabolites showing measurement order effects were corrected by subtracting the average of the corresponding cluster. The metabolite concentrations were then recalculated to the original magnitude scale. After this procedure, a total of 1,405 peaks remained. The intensity of these peaks was normalized across samples using upper-quartile normalization separately in each mode.

To eliminate metabolites with concentrations affected by a postmortem delay, we compared metabolite concentrations in macaque PFC samples collected from 42 individuals of different ages with a short PMD (post-mortem delay) of less than 20 minutes after death and with a prolonged PMD of 5-6 hours. For each metabolite, we assessed the difference between the concentration values in samples obtained after prolonged PMD and the concentration values interpolated at the corresponding age from a spline curve fitted to the concentration values measured in samples with a short PMD with four degrees of freedom. Metabolites with concentration values in samples with prolonged PMD falling outside the

95% confidence interval of spline curve predictions were excluded from further analysis. This procedure excluded 39 (2.8%) of the 1,405 metabolic peaks (Figure 3.1).

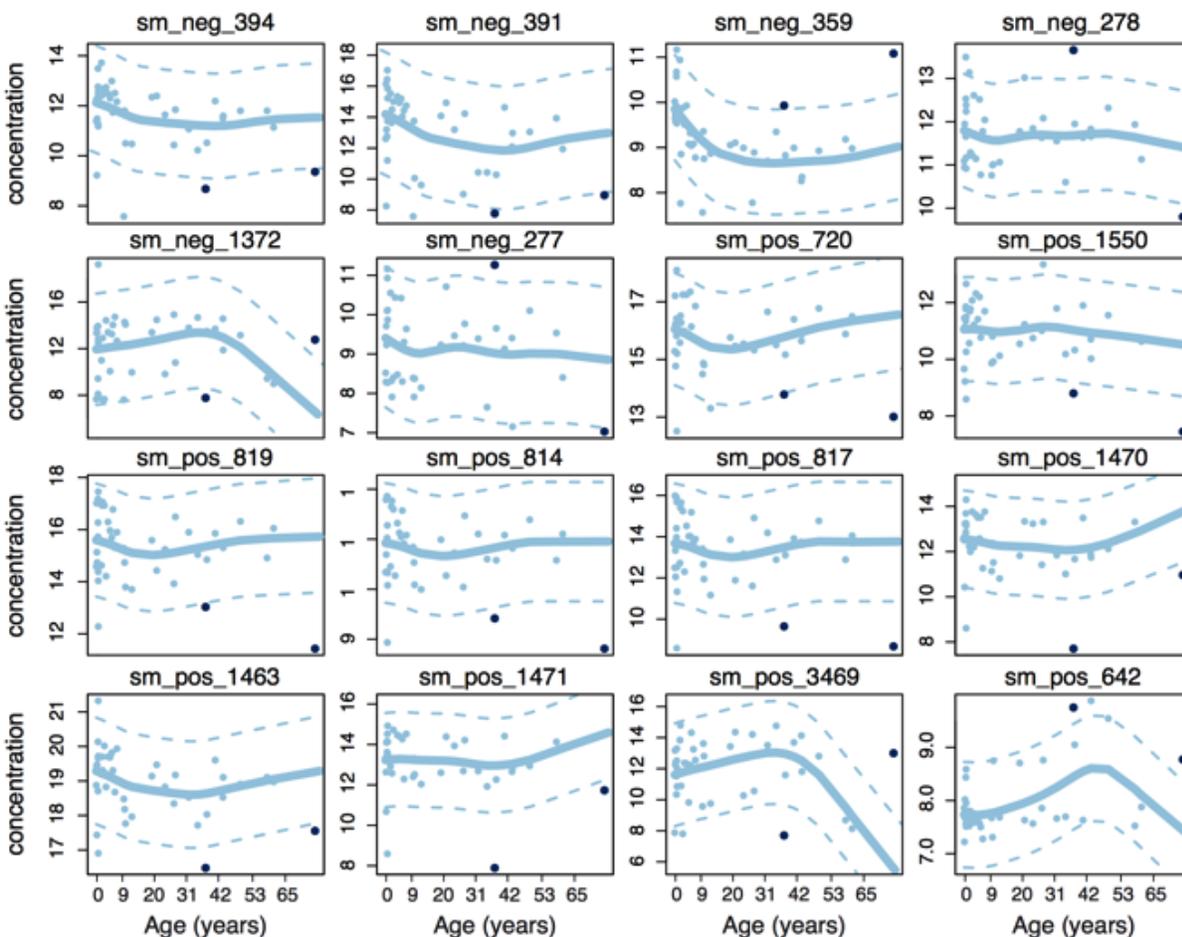


Figure 3.1: Concentration profiles of 16 unannotated metabolites, out of 39 metabolites classified as affected by postmortem delay in our analysis. The points indicate metabolite concentrations in the PFC of macaque individuals with short postmortem delays (<20 minutes, light blue) and two macaque individuals with long postmortem delay (5-6 hours, dark blue). The lines are the spline curves fitted to the data points with short postmortem delay individuals with four degrees of freedom. The dashed lines correspond to 95% confidence interval of spline curve predictions.

3.1.5 Statistical analysis

The resulting 1,366 peaks were used to explore relationships among samples with a multidimensional scaling algorithm (MDS) (Figure 3.3 B and 3.5 A).

To identify metabolites with concentration differences between ASD samples and unaffected controls, we used an analysis of covariance (ANCOVA) as described in [Somel et al., 2009, Liu et al., 2016]. 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 R^2 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.

To identify patterns of age-related concentration differences for ASD-related metabolites, we performed hierarchical clustering with 1 - Pearson correlation coefficient as the distance measure using both ASD and control samples. We used complete-linkage method of hierarchical clustering and cut the tree at 4 clusters (Figure 3.3 C). Segregation to a large number of clusters did not reveal novel patterns and yielded clusters with low numbers of metabolites ($n < 20$).

To test the overrepresentation of ASD-related metabolites in metabolic pathways, we performed a pathway enrichment analysis using R package clusterProfiler [Yu et al., 2012] based on the enzymes that were directly linked to these metabolites according to the KEGG database annotation [Kanehisa et al., 2017]. Enzymes directly linked to metabolites are enzymes, which either process or use as co-factor the following metabolites. Those interactions were directly taken from the KEGG database API. Genes directly linked to all the metabolites detected in our study were used as the background. The results were corrected for multiple testing using BH correction (Figure 3.4 A).

To test whether the disease status can be predicted using metabolite concentration levels, we implemented logistic regression with l_1 regularization, which is a linear model with an additional penalty on its coefficients that make it possible to train the model while simultaneously performing feature selection. Because the feature selection showed substantial variability among different training subsets of the data, we performed stability selection as described in [Meinshausen and Buehlmann, 2008], which is a procedure based on multiple subsampling of the data. Briefly, the dataset was divided into the training and test sets and the model was trained using the training set with a fixed number of parameters ($C=100$). For each trained model, we tested the performance of the classifier using a test set. By combining the results of all the data subsampling combinations ($n=500$), we calculated the mean performance and, for each metabolite, the empirical probability of being incorporated into

the model with a non-zero coefficient (Figure 3.2). We then ranked the metabolites according to this empirical probability. Based on model performance assessment, we chose a cutoff of the top 200 metabolites to define the set of model predictors.

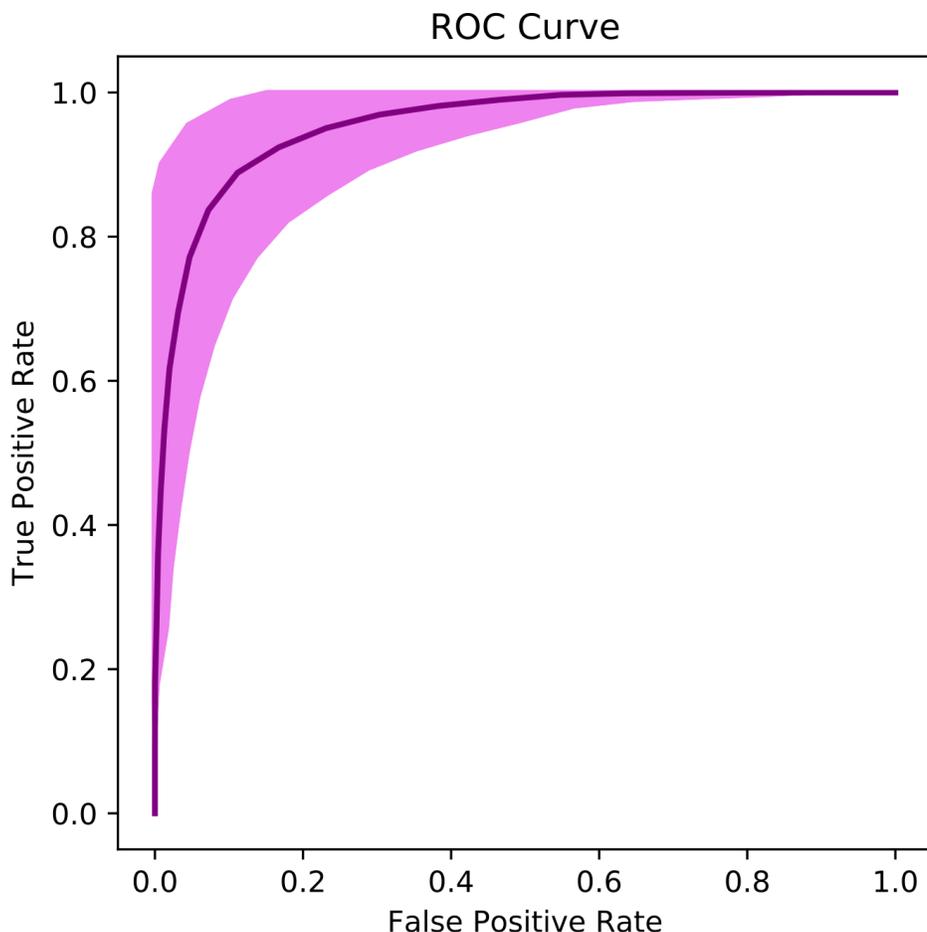


Figure 3.2: Area under the receiver operating characteristic curve (ROC AUC) calculated using logistic regression with optimal parameters performed 500 times on different subsets of the individuals. The solid line corresponds to average ROC AUC, while the area shows the standard deviation.

We used two sets of cutoffs (stringent and relaxed) to identify species-specific concentration differences. To identify differences using the stringent cutoff, we used the ANCOVA approach described in [Somel et al., 2009, Liu et al., 2016] for each species pair twice, using either species as a reference. If the test was significant for both human/chimpanzee and human/macaque pairs but not significant for the chimpanzee/macaque pair then the metabolite was classified as showing human-specific metabolic difference. Similarly, if the test was significant for both human/chimpanzee and chimpanzee/macaque pairs, but not significant for

the human/macaque pair then, the metabolite was classified as showing chimpanzee-specific metabolic difference. To conduct age alignment between species, we used the age scaling procedure described in [Li et al., 2017]. To identify differences using the relaxed cutoff, we calculated distances between species using macaque metabolite concentrations as a baseline. A metabolite was classified as showing human-specific metabolic difference if its human-macaque distance was larger than chimpanzee-macaque distance, and the direction of changes relative to the macaque coincided in humans and chimpanzees. Similarly, a metabolite was classified as showing a chimpanzee-specific metabolic difference if its chimpanzee-macaque distance was larger than human-macaque distance, and the direction of changes relative to macaque coincided between humans and chimpanzees. For each metabolite, the distance was calculated as the absolute difference between average z-transformed concentrations within species. To assess human-specificity in each module, we performed 1000 subsampling of the same number of metabolites and calculated the ratio of human-specific to chimpanzee-specific metabolic differences for each subsampling (Figure 3.5 B).

We tested the overrepresentation of human-specific metabolic differences defined using stringent criteria in KEGG pathways using R package clusterProfiler [Yu et al., 2012] as described above. To test the significance of the overlap between ASD-related and human-specific pathways, we performed Fisher’s exact test (Figure 3.5 C).

To compare ASD-related metabolites with genes differentially expressed in ASD, we estimated gene expression levels using published dataset deposited in the Gene Expression Omnibus (GEO) under accession number GSE28521 [Voineagu et al., 2011]. The dataset contains transcriptome data for 19 ASD samples and 17 control samples from three brain regions: the cerebellum, the frontal cortex, and the temporal cortex. To test whether genes linked to ASD-related metabolites tended to show an excess of expression levels particular to ASD, we calculated the proportion of genes with ASD-dependent expression linked to each metabolite. Genes showing absolute \log_2 fold change between the average concentrations of ASD samples and control samples greater than 0.2 were defined as genes with ASD-dependent expression in this analysis (Figure 3.4 E).

To compare concentrations of metabolites detected in our study with a previously published dataset of metabolite concentrations in humans, chimpanzees and macaques [Fu et al., 2011], we matched metabolites between the datasets using annotation. We then calculated \log_2 fold changes between the average metabolite concentrations in humans and the average metabolite concentrations in macaques (Figure 3.5 D). To test whether \log_2 fold change values agreed between datasets, we calculated the Pearson correlation coefficient and performed Fisher’s exact test.

3.2 Results

3.2.1 Metabolic data description

We searched for metabolic alterations in autistic brains by measuring concentrations of polar low molecular weight compounds in the postmortem prefrontal cortex (PFC) samples of 32 ASD patients and 40 control individuals. The control and ASD samples were matched for sex and sample quality, which was estimated using RNA integrity levels. Each group covered a broad age range: 2-60 years for ASD patients and 0-61 years for controls. All samples were represented by the cortical gray matter.

The measurements, conducted in a random order using liquid chromatography coupled with mass spectrometry (LC-MS) in positive and negative ionization modes, yielded 4,065 and 1,685 distinct MS peaks representing polar compounds (metabolites) with molecular weights below 2,000Da. Among them, 801 and 209 metabolites were computationally annotated using probabilistic matching to the Human Metabolome Database HMDB [Wishart et al., 2012] and the LIPID MAPS Structure Database (LMSD) [Fahy et al., 2009]. The removal of metabolites with concentrations influenced by the measurement order, experimental batch effects, and post-mortem delay yielded 1,366 confounder-free metabolic peaks detected in both ionization mode (Figure 3.3 A). Multidimensional scaling (MDS) of the ASD and control samples using normalized abundance of these 1,366 metabolites revealed a separation of very young individuals from the rest (Figure 3.3 B). Correspondingly, age explained 25% of the total metabolic variation among samples, while ASD accounted for 10%, postmortem intervals (PMI) accounted for 8%, and other factors such as sex, sample quality, and ethnicity accounted for less than 5% each.

3.2.2 Metabolic changes in autism

Of the 1,366 detected metabolites, 202 (15%) showed significant concentration differences between autism and control samples (autism-related metabolites, ANCOVA, BH-corrected $p < 0.05$). Unsupervised clustering of the temporal concentration profiles of these 202 metabolites revealed four modules (Figure 3.3 C). Genes linked to the 202 autism-related metabolites based on KEGG annotation were significantly overrepresented in a total of 16 KEGG (the Kyoto Encyclopedia of Genes and Genomes) pathways [Kanehisa et al., 2017] (hypergeometric test, BH-corrected $p < 0.05$, Figure 3.4 A).

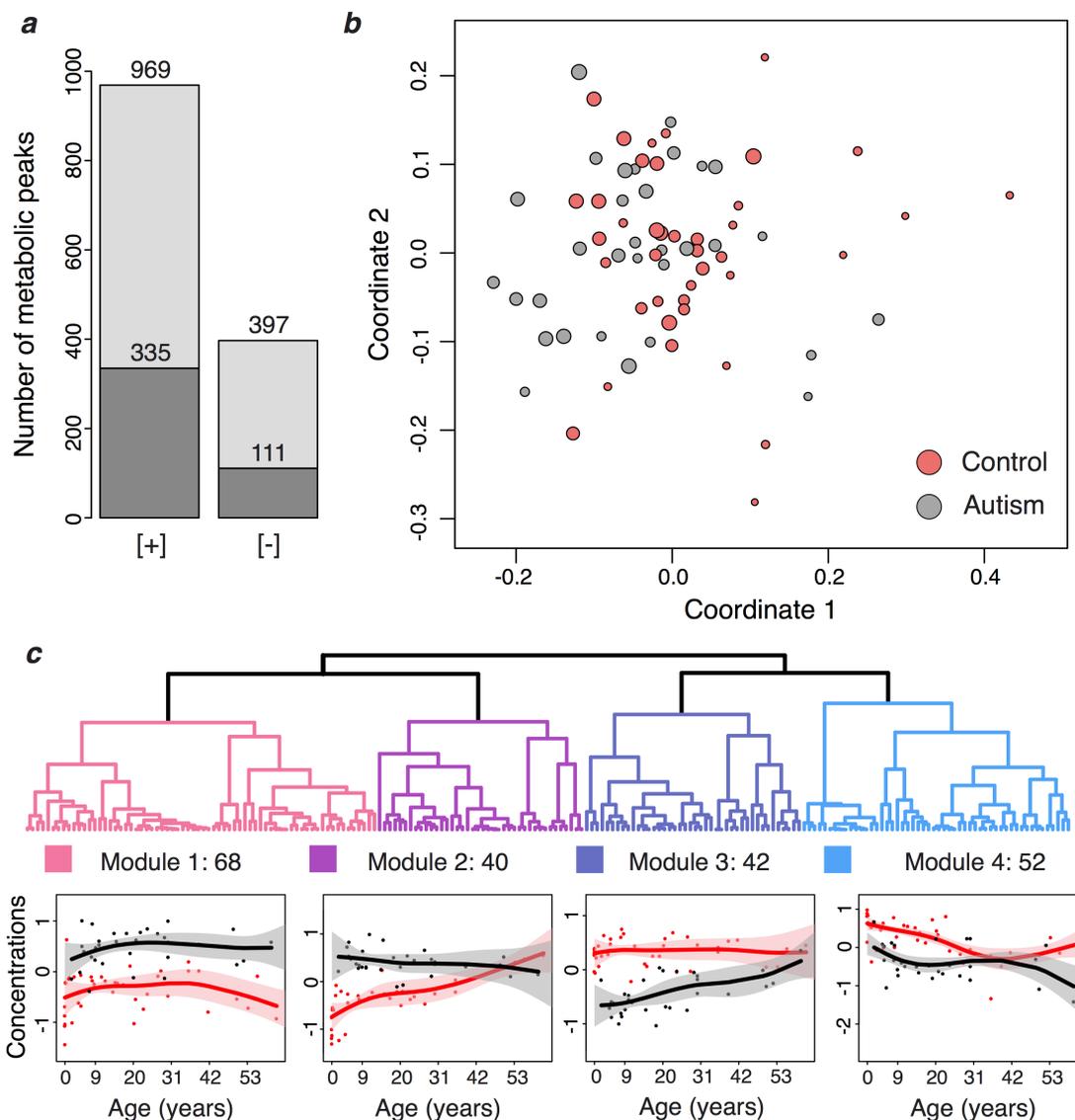


Figure 3.3: Identification of metabolic differences in the prefrontal cortex of autism patients. (a) The number of metabolic peaks detected by LC-MS procedure in positive ([+]) and negative ([-]) ionization modes, after removal of peaks affected by confounding factors. Darker shades indicate lipid peaks annotated using probabilistic matching to the LIPID MAPS database and HMDB. (b) The relationship among individuals plotted as the first two dimensions of the multidimensional scaling (MDS) procedure based on concentrations of 1,366 metabolites. Circles represent individual samples. Colors represent autism patients (gray) and control individuals (red). Size of the circles is proportional to the individual's age (smaller circles correspond to younger age). (c) Hierarchical clustering dendrogram based on concentrations of 202 autism-related metabolites and the concentration patterns in the four cluster modules. The metabolite concentration levels within each module were standardized to mean = 0 and standard deviation = 1. Points represent mean concentrations in each individual (red - controls; black - autism). Lines show cubic spline curves fitted to the data. Pink and gray shaded areas show one standard deviation of the curve estimates.

Notably, 10 of the 16 pathways were identified at similar significance level using 42 metabolites of the module 3 characterized by the concentration decrease in autism, especially early in life. These 10 pathways included glutathione metabolism, the pathway showing the strongest enrichment signal both globally and within module 3. Notably, the glutathione metabolism pathway contained not only significant concentration differences of glutathione and linked metabolites, but also genetic variants reported to be associated with autism in enzymes catalyzing the corresponding reactions (Figure 3.4 B,C). Module 4 containing 52 metabolites was significantly enriched in four of the 16 pathways, including strong enrichment in purine and pyrimidine metabolism pathways. By contrast, modules 1 and 2 characterized by elevated concentrations in autism were not substantially overrepresented within the 16 pathways, but were enriched in the four additional module-specific pathways including amino acid and nicotinamide metabolism.

3.2.3 Metabolic predictors of autism

Metabolite concentration differences between ASD cases and unaffected controls might allow for the classification of autistic brain samples as a distinct group using a machine-learning algorithms. To test this, we constructed a predictive model based on logistic regression with lasso regularization to assign each sample to the ASD or control group using the metabolite concentrations. We then performed stability selection [Meinshausen and Buehlmann, 2008], a procedure based on subsampling, to rank the metabolites' contribution to the model and to assess the accuracy. Remarkably, the model distinguished ASD and control cases with more than 95% accuracy, estimated as the area under the ROC curve (ROC AUC), which corresponds to the area under the curve mapping the true positive rate (sensitivity) to the false positive rate (1-specificity) for different discrimination thresholds (Figure 3.4 D). Thus, the AUC ROC provides a measure of the diagnostic ability of the classifier.

The model's metabolic predictors overlapped significantly with metabolites showing significant concentration differences in ASD in the ANCOVA analysis (Fisher test, $p < 0.001$). Consistently, KEGG pathways enriched in the top 200 metabolic predictors were in good agreement with pathways obtained in the ANCOVA analysis, with glutathione metabolism and purine metabolism occupying the top positions for each of the two methods (Figure 3.4 A).

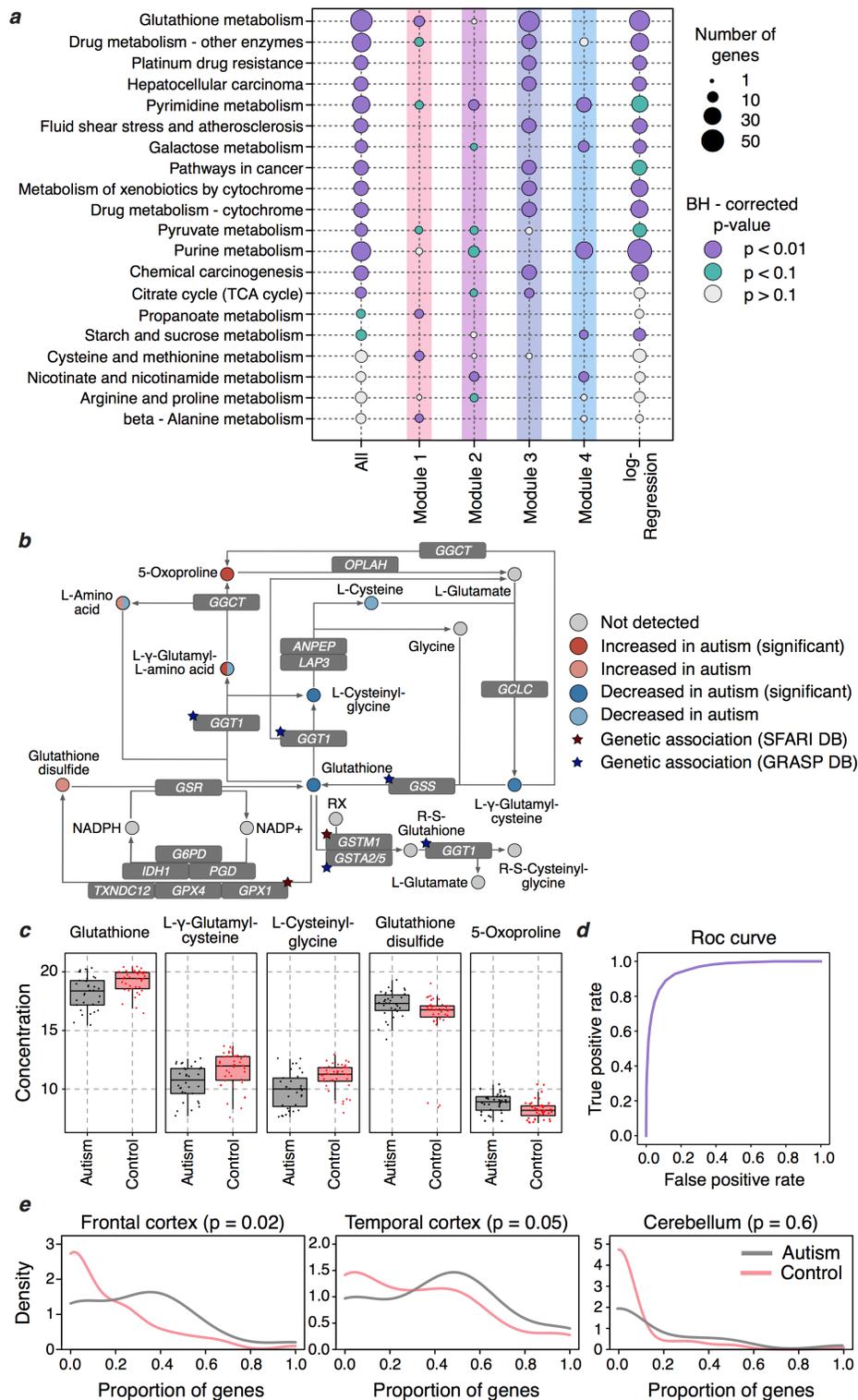


Figure 3.4: Characterization of metabolic differences identified in the prefrontal cortex of autism patients. (a) Summary of top functional pathways enriched in genes linked to metabolites represented in different categories using KEGG annotation. The categories include: all 202 autism-related metabolites identified using ANCOVA (ANCOVA); autism-related metabolites within each module (Module 1-4); and autism metabolic predictors

Figure 3.4: identified using logistic regression (log-Regression). Size of circles is proportional to the number of genes within the pathway linked to metabolites in a given category (smaller circles correspond to a smaller number of genes). The color of circles indicates BH-corrected enrichment p-value. (b) Simplified schematic representation of glutathione pathway based on the KEGG annotation. Circles represent metabolites. Circle colors indicate direction and significance of the difference. The double coloring of L- γ -glutamyl-L-amino acid and L-amino acid represent concentration changes of different compounds that fall under this biochemical annotation. Stars mark genes containing genetic variants associated with the ASD according to SFARI and GRASP databases. (c) The concentration of five metabolites from the glutathione metabolism pathway showing a differential concentration in ASD: glutathione, L- γ -glutamylcysteine, L-cysteinylglycine, glutathione disulfide, and 5-oxoproline. Boxes show the first and the third quartiles and the median of the data, the whiskers extend to the minimum and maximum data values located within 1.5 interquartile range from the box. Dots indicate actual concentration values for individual samples. Colors represent autism patients (gray) and control individuals (red). (d) Average area under the receiver operating characteristic curve (ROC AUC) calculated using logistic regression with optimal parameters performed 500 times on different subsets of the individuals. (e) Distributions of the proportions of genes showing expression difference in autism ($|\log_2 \text{fold change}| > 0.2$) among genes linked to each of 202 autism-related metabolites (gray curve) and the other metabolites detected in our study (red curve). Expression data was taken from [Voineagu et al., 2011].

3.2.4 Gene expression associated with metabolic changes in autism

To test whether the metabolic differences between ASD and control cases might be caused by gene expression differences, we examined data from three brain regions measured in 19 ASD patients and 17 controls, including 15 patients and 5 controls from the present study [Voineagu et al., 2011]. In two cortical regions, but not in the cerebellum, genes linked to ASD-related metabolites indeed showed significantly more expression differences between ASD patients and controls than genes linked to metabolites showing no concentration difference in ASD (KS-test, $p < 0.05$, Figure 3.4 E).

3.2.5 Evolution of metabolic changes in autism

Previous work linked gene expression differences in ASD to developmental gene expression features unique to humans [Liu et al., 2016]. To assess this link at the metabolite level, we analyzed metabolite concentrations in the PFC of 40 chimpanzees with ages between 0 and 42 years, and 40 rhesus macaques with ages between 14 weeks post-conception and 21 years. The chimpanzee and macaque samples were measured together with human control and ASD samples in a random order. A computational analysis of nonhuman primates,

conducted in parallel with the human samples, yielded concentrations in chimpanzee and macaque samples for 1,366 metabolites confidently detected in human dataset. A multidimensional scaling (MDS) analysis based on the concentrations of these metabolites revealed predominant sample separation according to age and species identity (Figure 3.5 A).

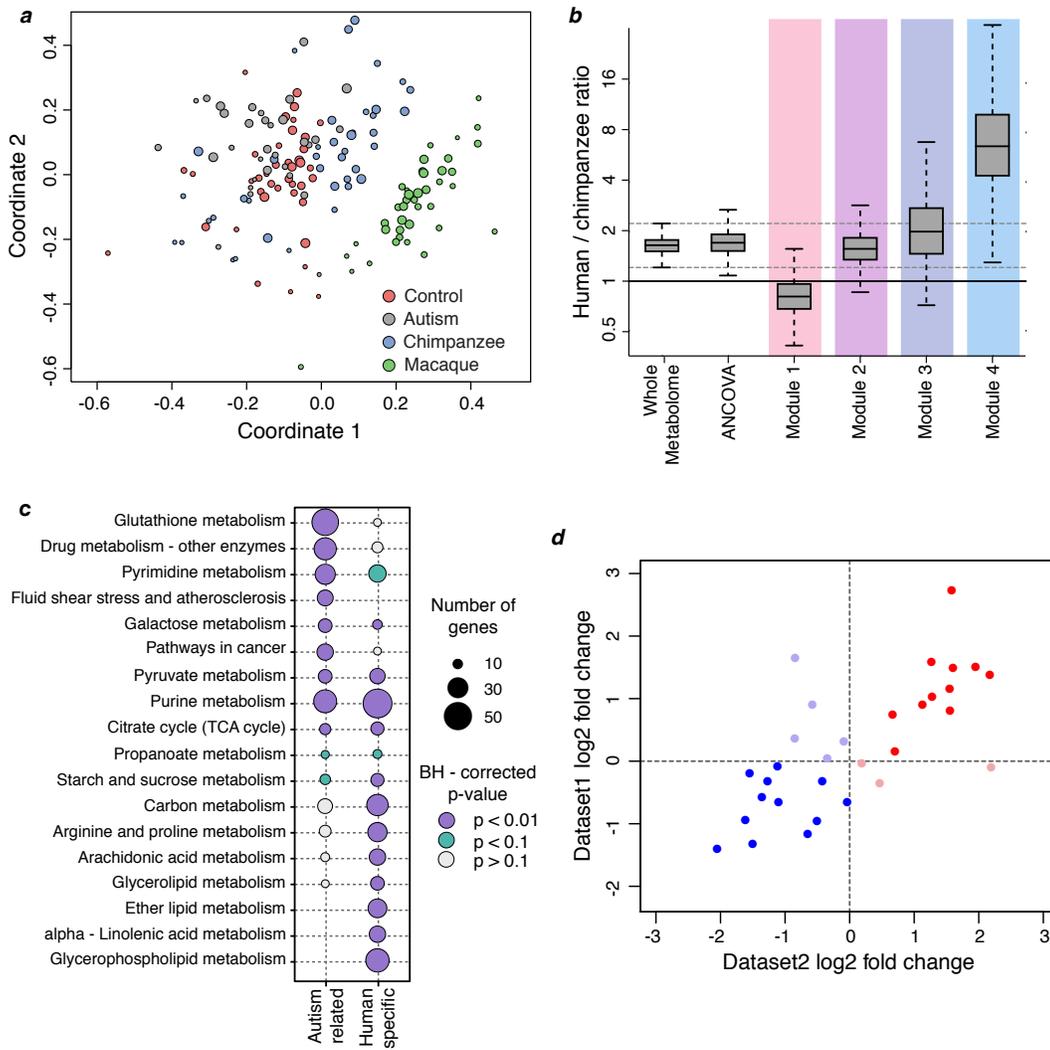


Figure 3.5: Evolution of metabolic differences identified in the prefrontal cortex of autism patients. (a) The relationship among species and individuals plotted as the first two dimensions of the multidimensional scaling (MDS) procedure based on concentrations of 1,366 metabolites. Circles represent individual samples. Colors represent: autism patients (gray), control human individuals (red), chimpanzees (blue), and macaques (green). Size of the circles is proportional to the individual's age (smaller circles correspond to younger age). (b) The ratio of human-specific and chimpanzee-specific metabolites represented in different categories: all 1,366 detected metabolites (Detected Metabolome); all 202 autism-related metabolites identified using ANCOVA (ANCOVA); and autism-related metabolites within

Figure 3.5: each module (Module 1-4). Boxes show the first and the third quartiles and the median of the data, the whiskers extend to the minimum and maximum data values located within 1.5 interquartile range from the box. Dashed gray lines indicate the interwhisker range of the detected metabolites. (c) Summary of top functional pathways enriched in genes linked to metabolites represented in two categories using KEGG annotation. The categories include: all 202 autism-related metabolites identified using ANCOVA (autism-related) and human-specific metabolites (human-specific). Size of circles is proportional to the number of genes within the pathway linked to metabolites in a given category (smaller circles correspond to a smaller number of genes). The color of circles indicates BH-corrected enrichment p-value. (d) Metabolite concentration differences between humans and macaques measured using our data (Dataset 1) and a published dataset (Dataset 2) [Fu et al., 2011]. The concentration differences were calculated as \log_2 -transformed differences between the average concentration values within each species. Dots represent individual metabolites detected in both datasets (N=31). Colors indicate plot quadrants.

The identification of significant metabolite concentration differences and their placement on the evolutionary lineages (ANCOVA, BH-corrected $p < 0.05$) revealed 170 human-specific and 55 chimpanzee-specific differences defined using strict criteria, and 756 human-specific and 410 chimpanzee-specific differences defined using relaxed criteria. Human-specific concentration differences detected in this study agreed well with the differences calculated using the published metabolome dataset (Figure 3.5 D, Pearson correlation $R=0.71$, $p < 0.01$) [Fu et al., 2011].

Genes linked to human-specific metabolites were significantly overrepresented in a total of 27 KEGG pathways (hypergeometric test, BH-corrected $p < 0.05$, Figure 3.4 F). Notably, these pathways overlapped significantly with pathways enriched in ASD-related metabolic differences, and included purine, pyrimidine, and pyruvate metabolism (Fisher-test, $p < 0.01$, Figure 3.5 C).

While the ratio of human-specific metabolic differences to the chimpanzee-specific ones among 202 ASD-related metabolites coincided with the one for all detected metabolites, their distribution among the four ASD modules was not uniform. Module 1 contained fewer human-specific metabolic differences compared to the average, while module 4 contained approximately five times more (Wilcoxon test, $p < 0.01$, Figure 3.5 B). Notably, module 4 was particularly enriched in metabolic pathways overlapping between ASD-related and human-specific differences, including purine and pyrimidine metabolism (Fisher test, $p < 0.1$).

3.3 Discussion

In this study we show that the metabolite composition of the gray matter of the prefrontal cortex differs substantially between ASD patients and healthy controls. Despite the moderate statistical power of the study, as many as 15% of metabolites present in the prefrontal cortex alter their concentrations significantly in ASD. These metabolites form specific age-dependent concentration patterns and cluster in specific KEGG pathways.

It is noteworthy that, of the 16 pathways altered in the prefrontal cortex of ASD patients based on ANCOVA results, 10 were reported in studies analyzing metabolite concentrations of urine and blood samples, which is significantly more than expected by chance (Fisher test, p -value <0.05). These pathways include glutathione metabolism, purine metabolism, pyruvate metabolism, propanoate metabolism, TCA cycle, galactose metabolism, starch and sucrose metabolism, nicotinate and nicotinamide metabolism, cysteine and methionine metabolism, and arginine and proline metabolism. Metabolic differences clustering in these 10 pathways were all reported in ASD patients' urine [Yap et al., 2010, Ming et al., 2012, Emond et al., 2013, Mavel et al., 2013, Cozzolino et al., 2014, Nadal-Desbarats et al., 2014, Kałuzna-Czaplińska et al., 2014, Noto et al., 2014, Diémé et al., 2015, Gevi et al., 2016, Lussu et al., 2017, Bitar et al., 2018]. Metabolic differences clustering in the TCA cycle, glutathione metabolism, and pyruvate metabolism pathways were further reported in ASD patients' blood [Rossignol and Frye, 2012, West et al., 2014, Wang et al., 2016, Laszlo et al., 1994, James et al., 2004]. Concentration differences of individual metabolites, such as glutathione and cysteine concentration decrease, as well as a glutathione disulfide and adenosine concentration increase in ASD patients identified in our study (Figure 3.4 C), were similarly reported in blood [James et al., 2004]. Furthermore, the same direction and magnitude of the concentration differences between ASD patients and controls were observed for glutathione and glutathione disulfide in the cerebellum and temporal cortex [Rose et al., 2012]. Similarly, concentration differences of 3-methoxytyramine and 5,6-dihydrouridine that increased in the cerebellum of ASD patients [Graham et al., 2016] were detected and reproduced in our dataset. At the gene expression level, the differences reported in the neocortex of ASD patients [Voineagu et al., 2011] were greater for genes linked to ASD-related metabolites identified in our study compared to genes linked to the other detected metabolites. Taken together, these observations confirm the previous results and indicate a general agreement between ASD-related metabolic differences detected in the brain and differences detected in the blood and urine analysis. Purine metabolism is particularly interesting, as purinergic signaling is involved in neurodevelopment processes, including cell proliferation, differentiation and neuron-glia cross-talk [Burnstock, 2011, Burnstock et al., 2011]. Moreover, purinergic

signaling was shown to be altered in multiple psychiatric disorders, including ASD [Cheffer et al., 2018].

Our analysis further yielded a number of novel observations. First, we uncovered novel metabolic differences complementary to previously reported ones. For instance, within the glutathione pathway, we show that in addition to a glutathione concentration decrease, two other metabolites, L-cysteinylglycine and L- γ -glutamyl-L-cysteine, display concentration differences in ASD. Further, enzymes catalyzing reactions involving glutathione, L-cysteinylglycine, and L- γ -glutamyl-L-cysteine contain genetic variants previously linked to ASD [Ming et al., 2010, Buyske et al., 2006, Anney et al., 2010].

Second, we identified a number of metabolic pathways that were not previously linked to ASD, including pyrimidine metabolism, β -alanine metabolism, and three pathways related to drug metabolism by cytochromes. Metabolic processes in these pathways are connected to the reported ones. For instance, changes in pyrimidine metabolism might be linked to purine metabolism alterations, as both pathways involve the essential precursors for RNA and DNA synthesis and are well interconnected. Furthermore, pyrimidine nucleotides are involved in polysaccharide and phospholipid synthesis, detoxification processes, as well as proteins' and lipids' glycosylation [Loffler et al., 2005]. Notably, functional deficiency of pyrimidine pathway genes, such as uridine monophosphate synthase (*UMPS*), dihydropyrimidine dehydrogenase (*DPYD*), Dihydropyrimidinase (*DPYS*), beta-ureidopropionase 1 (*UPB1*), leads to neurological aberrations, including autism-like features in case of DPD deficiency (*DPYD* gene deficiency) [Micheli et al., 2011]. Metabolic changes in beta-alanine pathway might be linked to pyrimidine metabolism pathway, as the main sources of β -alanine include catabolism of cytosine and uracil. Moreover, β -alanine concentration is decreased in DPD (*DPYD* gene deficiency), DHP (*DPYS* gene deficiency), BUP-1 (*UPB1* gene deficiency) induced pyrimidine pathway deficiencies [Micheli et al., 2011]. In addition, key enzymes participating in cytochrome mediated drug metabolism are also involved in glutathione metabolism pathway.

Third, we show that machine learning algorithms could accurately separate ASD patients from unaffected controls using brain metabolite concentrations. Notably, metabolites used by the predictive model cluster in 14 of the 16 pathways were discovered using ASD-related metabolites identified by ANCOVA.

Fourth, by conducting brain metabolome measurements in nonhuman primates, we show that ASD-related metabolites falling within module 4 contain an almost 8-fold excess of metabolic differences unique to humans, compared to the chimpanzee-specific differences. Accordingly, a number of pathways enriched in ASD-related metabolic differences also show an excess of human-specific metabolic differences. These pathways include purine, pyru-

vate metabolism, TCA cycle, and galactose metabolism. Moreover, these results agree with previous study results obtained using GC-MS, which reported an excess of human-specific metabolite concentration differences [Fu et al., 2011]. Taken together, these findings align well with the hypothesis postulating disruption of recently evolved cognitive mechanisms underlying communication and socialization in ASD.

More generally, our results indicate the presence of substantial brain metabolism alteration in ASD. Remarkably, many of these alterations, both at the pathway level and at the level of individual metabolites, coincide with the differences reported in urine and blood samples. This finding underscores the potential of metabolic studies conducted in non-nervous tissues. At the same time, our brain tissue analysis reveals additional metabolic differences, contributing to a more systematic understanding of the disease.

Chapter 4

Lipidome alterations in human prefrontal cortex in cognitive disorders

In this chapter, we investigated involvement of lipids in brain disorders, for this purpose we conducted mass spectrometry-based untargeted lipidome profiling in prefrontal cortex of 403 cognitively healthy humans with ages spanning the entire lifespan, as well as 17 autism (ASD), 27 schizophrenia (SZ) and five Down syndrome (DS) patients. Of the 5,024 hydrophobic compounds (lipids) detected in our dataset, 10.8%, 10.4% and 2.6% of detected lipids alter their concentrations significantly in DS, SZ and ASD, respectively. Lipids associated with the three disorders have convergent enrichment in glycerophospholipid metabolism and endocannabinoid signaling pathways. Moreover those results were supported by an independent dataset, which contained 40 cognitively healthy humans and 33 ASD patients.

4.1 Materials and Methods

4.1.1 Samples

Samples from cognitively unaffected human controls were collected from the NICHD Brain and Tissue Bank for Developmental Disorders and the Maryland Psychiatric Research Center at the University of Maryland, the Maryland Brain Collection Center, the Netherlands Brain Bank, and the Chinese Brain Bank Center (CBBC, <http://cbbc.scuec.edu.cn>, Wuhan, China). Samples from patients with Down syndrome were collected from the Netherlands Brain Bank. Samples from patients with autism or schizophrenia were collected from NICHD Brain and Tissue Bank for Developmental Disorders at the University of Maryland, the Harvard Brain Tissue Resource Center and the Maryland Psychiatric Research Center. Brain

tissue from subjects with autism was obtained through the Autism Tissue Program [Brimacombe et al., 2007]. Diagnosis of schizophrenia is based on DSM IV-based criteria and was provided by the participating brain banks. Written consents for the use of human tissues for research were obtained either from the donors or their next of kin. According to the protocol of the CBBC, specific permission for brain autopsy and use of the brain tissue for research purposes was given by the donors or their relatives. All tissue samples were shipped by the brain banks without accompanying personal identifier information. All human samples in this project were extracted from the prefrontal cortex (PFC), which were dissected from the anterior part of the superior frontal gyrus. The sample weights are 12.55 ± 1.65 mg. All samples were well preserved postmortem samples that had been stored at -80°C before RNA or lipid extraction.

4.1.2 MS sample preparation and measurements

Metabolites were extracted from 10-15 mg of frozen tissue, which was homogenized by a ball mill to a fine powder, as described elsewhere [Bozek et al., 2015]. In brief: the frozen tissue was transferred to cooled 2 ml round bottom microcentrifuge tubes and each sample was re-suspended in 1 ml of a -20°C methanol:methyl-tert butyl-ether (1/3 (vol/vol)) mixture, containing 1.5 μg of 1,2-diheptadecanoyl-sn-glycero-3-phosphocholine (Avanti Polar Lipids, 850360P). The samples were immediately vortexed before they were incubated for 10 min at 4°C on an orbital shaker. This step was followed by ultra-sonication in an ice-cooled bath-type sonicator for additional 10 min. To separate the organic from the aqueous phase 650 μl of a H_2O :methanol mix (3/1 (vol/vol)) was added to the homogenate, which was shortly vortexed before it was centrifuged for 5 min at 14,000 g. Finally, 500 μl of the upper MTBE phase, which contains the hydrophobic compounds (lipids), were sampled to a fresh 1.5 ml microcentrifuge tube. This aliquot can either be stored at -20°C for some weeks or immediately dried in a speed vacuum concentrator at room temperature.

Prior to analysis, the dried pellets were re-suspended in 400 μL acetonitrile:isopropanol (7/3 (vol/vol)), ultra-sonicated and centrifuged for 5 min at 14,000 g. The cleared supernatant was transferred to fresh glass vials and 2 μL of each sample was injected onto a C8 reverse phase column (100 mm x 2.1 mm x 1.7 μm particles, Waters) using a UPLC system (Acquity, Waters, Manchester, UK). In addition to the individual samples, we prepared pooled samples, namely 10 μl of each sample was mixed. These pooled samples were measured after every 20th sample, providing us information on system performance including information on sensitivity, retention time consistency, sample reproducibility and compound stability.

The mobile phase for the chromatographic separation consisted of (Buffer A): 1% 1M ammonium acetate, 0.1% acetic acid in UPLC MS grade water (BioSolve, Valkenswaard, Netherlands), while Buffer B consisted of 1% 1M ammonium acetate, 0.1% acetic acid in acetonitrile/isopropanol (7/3 (vol/vol)), Biosolve). The flow rate of the UPLC system was set to 400 $\mu\text{L}/\text{min}$. The gradient was: 1 min isocratic flow at 45% A, 3 min linear gradient from 45% to 25% A, 8 min linear gradient from 25% to 11% A, 3 min linear gradient from 11% to 1% A. After cleaning the column for 4.5 min at 1% A, the buffer was set back to 45% A, and the column was re-equilibrated for 4.5 min, resulting in a final run time of 24 min per sample.

The mass spectra were acquired using an Orbitrap-type mass spectrometer (Orbitrap-XL, Thermo-Fisher, Bremen, Germany). The spectra were recorded using full scan mode, covering a mass range from 100-1500 m/z . The resolution was set to 60,000 with 2 scans per second, restricting the maximum iontime to 100 ms. The samples were injected using the heated electrospray ionization source (HESI) and the capillary voltage was set to 3.5 kV in positive and negative ionization mode. The sheath gas flow value was set to 40, while the auxiliary gas flow was set to 20. The capillary temperature was set to 200°C, while the drying gas in the heated electro spray source was set to 350°C. The skimmer voltage was set to 20 V while the tube lens was set to a value of 140V. The spectra were recorded from 0 to 20 min of the UPLC gradients. Sample randomization was performed twice: before the lipid extraction and before the mass spectrometry measurements. Persons performing lipid extraction and lipidome measurement were unaware of sample information, including age and diagnostic status. The correlation between the measurement order and sample age were not significant (positive ionization mode: $\rho=0.006$, $P>0.8$; negative ionization mode: $\rho=-0.002$, $P>0.9$). Similarly, there was no significant relationship between measurement order and health status or ethnicity (Wilcoxon rank sum test, nominal $P>0.3$). The MS preparation and measurement procedure for autistic samples in DS2 has been described elsewhere [Li et al., 2017].

4.1.3 Lipidome data preprocessing

MS lipids were extracted and aligned across samples using Progenesis QI software (Version 2.3, Nonlinear Dynamics, Newcastle upon Tyne, UK) according to the vendor description in order to perform peak peaking. Next the software automatically detects the different isotope peaks, clustered them together, and reported summed intensity as the monoisotopic mass retention time feature. The resulting features were next searched against LIPID MAPS with a mass tolerance of 10 ppm, allowing $[M+H]$, $[M+NH_4]$, $[M+Na]$, $[M-H_2O+H]$ modifications as

possible adducts in the positive ionization mode, and $[M-H]$ and $[M+oAC-H]$ modifications in the negative ionization mode. Finally, only the reliably detected lipids satisfying the following criteria were retained: a) retention time (RT) ≥ 0.6 min; b) detected in at least 90% of pooled samples and at least 80% of non-pooled samples. Quantile normalization was applied to the \log_{10} -transformed concentration of reliably detected lipids. The batch effect correction was done based on a linear regression model.

Data preprocessing procedure for lipid concentrations measured in DS2 has been described elsewhere [Li et al., 2017]. The batch effect correction was done using the fitted support vector regression model with a Gaussian kernel, considering the concentration of each lipid as a function of the measuring order of the samples. An upper quartile normalization was used to normalize the lipid concentration measurements.

4.1.4 Lipid annotation and enrichment analysis

Lipid annotations were performed using mass search with a tolerance of 10 ppm against the LIPID MAPS database annotation [Fahy et al., 2009], using the list of adducts described elsewhere [Bozek et al., 2015]. Lipid classes were assigned to the annotated lipids according to the LIPID MAPS database classification. Pathways were assigned according to the KEGG database annotation [Kanehisa et al., 2017]. Overrepresentations of lipid classes and pathways compared to random sampling from detected lipids with annotations were tested using one-sided Fisher’s exact tests and hypergeometric tests, respectively, followed by BH corrections. Significant enrichment was defined as BH-corrected $P < 0.05$.

4.1.5 Disorder-associated lipids identification

Disorder-associated analyses were conducted for patient samples of each neural disorder and matched control samples, respectively. To eliminate the confounding effects of ages, only the cognitively healthy samples with matched age ranges as the patient samples, i.e. 18 to 65 years old, were used. The Han Chinese cognitively healthy samples were also removed from the analysis to avoid confounding the effect of ethnicities. Two latent batches in the dataset were detected by hierarchical clustering (complete linkage), based on the pairwise distance between each pair of samples defined as $1 - \rho$, where ρ is the Pearson’s correlation coefficient between the two samples across concentrations of all the detected lipids. The latent batch effect on the lipidome was corrected using the linear model as that for the batch effect correction.

Based on the corrected lipid concentrations, the disorder-associated lipids were identified using Wilcoxon rank sum test to compare the lipid concentrations in patient samples of each disorder and the control samples, with nominal $P < 0.01$. Permutation of diagnostic status 1,000 times were used to estimate test significance.

For autism DS1 and DS2 analysis, we used methodology described in chapter 3.1.5.

4.1.6 Quantification of disorder effect on lipid concentration variance

To estimate the lipid concentration variance explained by the diagnostic status, we fit a linear model for each lipid, using lipid concentration as the response variable and diagnostic status as the predictor variable. The disorder effect was quantified as the ratio of the explained sum of squares (ESS) relative to the total sum of squares (TSS). To perform estimations with age, sex, PMI distributions and sample size equalized among disorders, we performed bootstrapping 100 times (sample size=5) and fit two models to each detected lipid. The null model considered the effect of age, sex and PMI on lipid concentration, while the full model further considered the effect of the disorder. The proportion of variance explained by diagnostic status was calculated as the difference of the residual of null model and that of the full model relative to the residual in the null model.

4.2 Results

4.2.1 The PFC lipidome composition is altered in common cognitive disorders

We measured the lipidome composition in PFC samples from 26 schizophrenia (SZ), 16 autism (ASD) and five Down syndrome (DS) patients aged 18 to 65 (Figure 4.1 A). The disorder samples were measured together with 403 samples from 396 cognitively healthy individuals in a random order.

For the 18-65 age interval, the diagnostic status explained the largest proportion of the total lipid concentration variation (6%, permutations, $P < 0.01$), compared to age (4%, permutations, $P < 0.01$) and PMI (1.7%, permutations, $P < 0.01$). Other factors, i.e. ethnicity, sample quality and sex did not affect lipidome composition significantly (permutations, $P > 0.15$). The effect of the diagnostic status was robust to correction for sample size, age, sex and PMI for all the three disorders (Wilcoxon rank sum test, Bonferroni-corrected $P < 0.0001$; Figure 4.2).

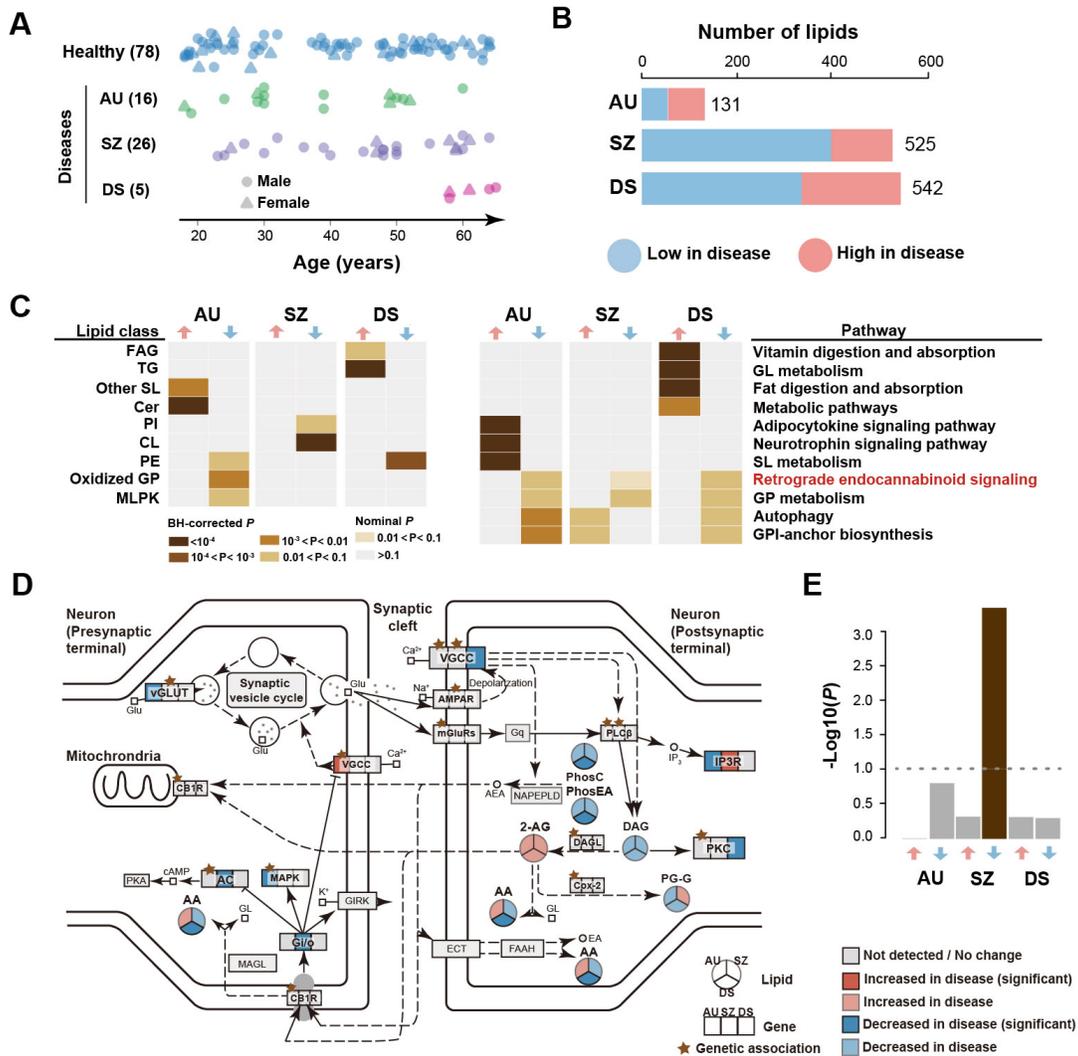


Figure 4.1: Disorder-associated (DA) lipidome changes in the human PFC [Yu et al., 2018]. (a) Age distribution of patients with autism (ASD, green), schizophrenia (SZ, purple), Down syndrome (DS, pink), and matched controls (blue). Each symbol represents an individual (circle males, triangle females). Numbers in brackets show numbers of individuals in each group. (b) Numbers of DA lipids for each disorder (red - higher concentration in disease, blue - lower concentration in disease). (c) Lipid classes (left) and KEGG pathways (right) enriched in DA lipids. Colors show P values of hypergeometric tests. FAG: fatty acyl glycosides, MLPK: macrolides and lactone polyketides, Cer: ceramides, OtherSL: other sphingolipids, PI: glycerophosphoinositols, PE: glycerophosphoethanolamines. (d) Schematic representation of DA genetic, gene expression and lipid concentration changes in retrograde endocannabinoid signaling pathway based on the KEGG annotation. Stars mark genes containing genetic variants linked to corresponding disease. Genes with SZ and ASD related genetic variants were retrieved from GRASP ($P < 0.05$) and SFARI respectively. Gene expression level changes were calculated using public data retrieved from GEO (ASD: GSE28521, SZ: GSE53978, DS: GSE5390). (e) Enrichment of DA genetic variants in genes linked to DA lipids. The y-axis shows the $-\log_{10}$ -transformed P value of the hypergeometric test.

Statistically, 10.8%, 10.4% and 2.6% of detected lipids altered their concentrations significantly in DS, SZ and ASD, respectively (disorder-associated (DA) lipids; Wilcoxon rank sum test, nominal $P < 0.01$; $N = 542$ for DS, $N = 525$ for SZ, $N = 131$ for ASD; permutations, $P < 0.01$ for SZ and DS, $P = 0.06$ for ASD; Figure 4.1 B). Additionally we demonstrated that there is a significant difference in gene expression between disease and control samples, represented by absolute values of \log_2 -transformed fold changes, for genes directly interacting with DA lipids based on KEGG annotation (Wilcoxon rank sum test, $P < 0.001$; Figure 4.2).

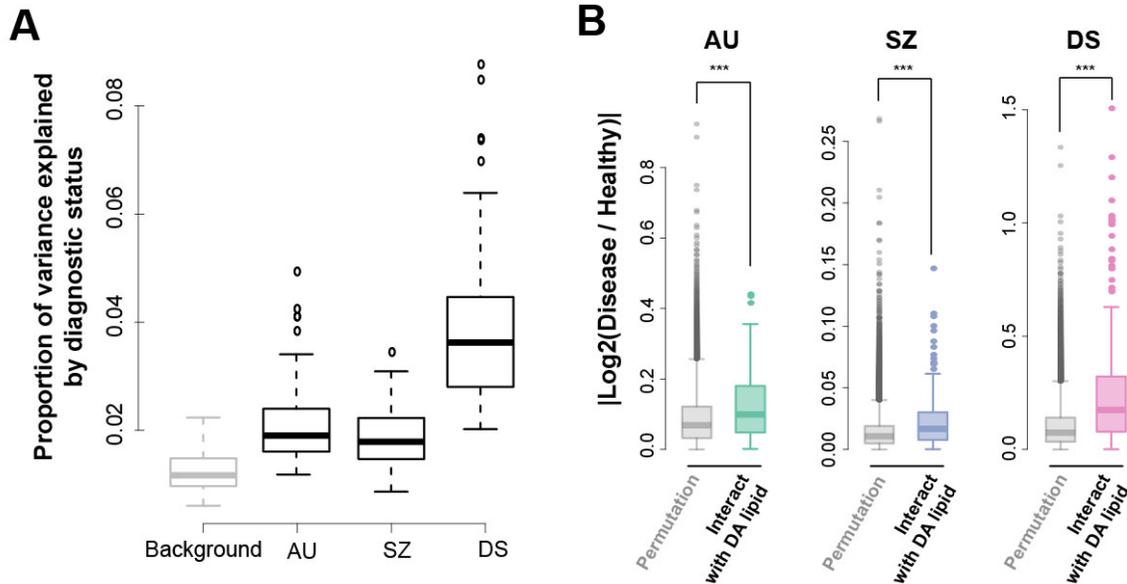


Figure 4.2: Lipidome and transcriptome alterations in the three cognitive diseases [Yu et al., 2018]. (a) Distributions of the average proportions of lipidome concentration variance explained by diagnostic status, based on bootstrapping over samples with sample size, age, sex, and PMI distributions equalized among three disorders 100 times. The background is estimated based on comparisons of randomly selected samples from the matched control samples to all matched control samples 100 times. (b) Distributions of gene expression differences between disease and control samples, represented by absolute values of \log_2 -transformed fold changes, for genes directly interacting with DA lipids based on KEGG annotation (colored boxes). The gray boxes show background distributions calculated by permuting disease and controls labels 1,000 times. Stars indicate significant differences between distributions in the two-sided Wilcoxon rank sum test (***: nominal $P < 0.001$). Gene expression differences were calculated using public datasets retrieved from GEO: ASD-GSE28521, SZ-GSE53978, DS-GSE5390.

Of the six disorder-associated lipid groups, sorted by disorders and concentration change directions, five were enriched in specific lipid classes and all six were enriched in functional pathways (one-sided Fisher's exact test, BH corrected $P < 0.1$; Figure 4.1 C). Notably, general

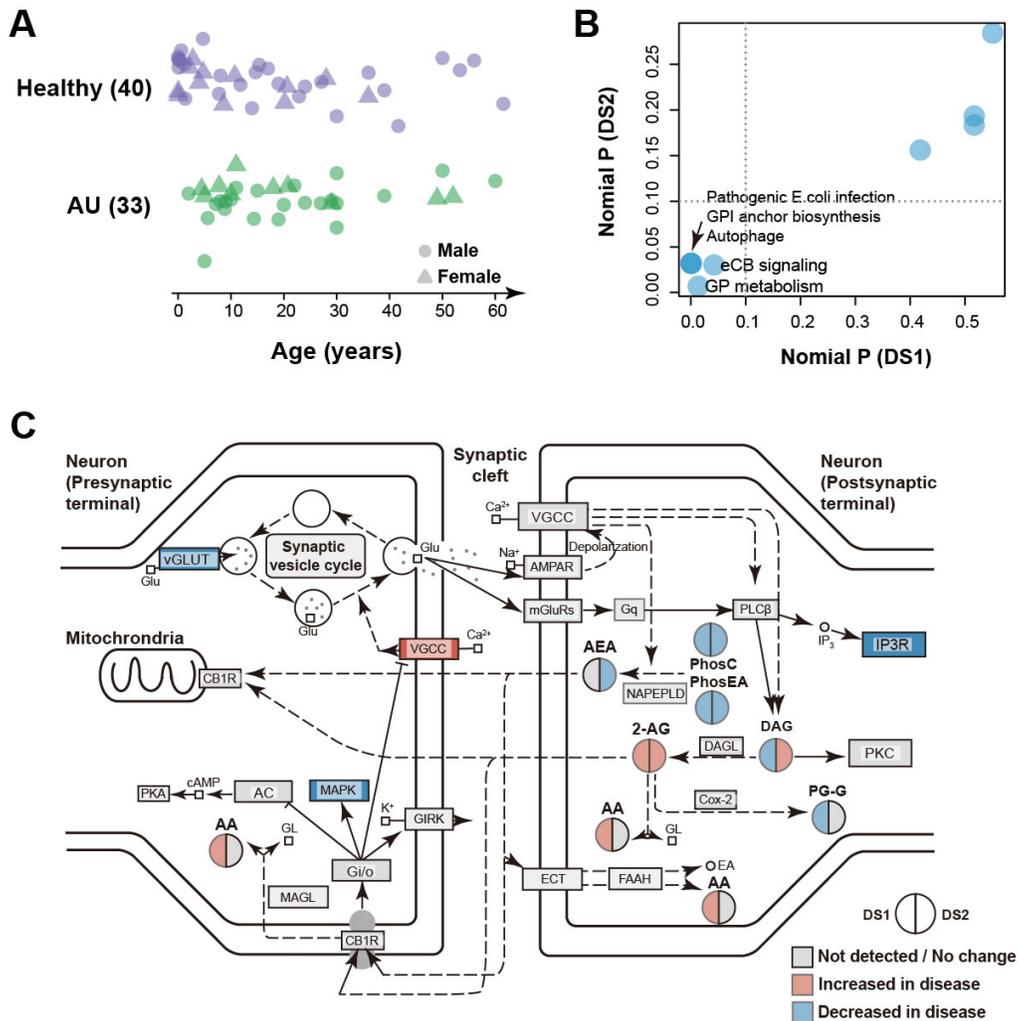


Figure 4.3: Autism-associated lipidome changes in DS2 [Yu et al., 2018]. (a) Age distribution of the PFC samples from matched controls (purple) and patients with autism (green) in DS2. Each symbol represents an individual (circle - male, triangle - female). Numbers in brackets show numbers of individuals for each group. (b) Nominal P values of hypergeometric test in pathway enrichment analysis calculated based on DS1 (x-axis) and DS2 (y-axis) lipids with decreased concentrations in autism samples identified independently in a respective dataset. Each circle represents one pathway. The black arrow marks three overlapping circles corresponding to three pathways with the same enrichment significance level, with names of these and other pathways enriched in both datasets (nominal $P < 0.1$) listed. (c) The schematic representation of autism-associated gene expression and lipid concentration changes in retrograde endocannabinoid (eCB) signaling pathway. The structure of the pathway is based on the KEGG annotation.

lipid concentration decrease in "retrograde endocannabinoid signaling" and "glycerophospholipid (GP) metabolism" pathways was shared among all three disorders. By contrast,

pathways enriched in lipids showing increased concentrations in disorders were particular to each disease (Figure 4.1 C, D).

To assess the validity of these results, we designed a replication experiment and generated an independently measured lipidome dataset containing 33 PFC samples of autism patients used in [Liu et al., 2016] and 40 matched controls included in DS2 processed in randomized order (autism DS2). The analysis, based on 9,058 lipids detected in this dataset, yielded consistent pathway enrichment results obtained using all 600 independently identified autism-associated lipids (one-sided Fisher's exact test, $P=0.02$), as well as 319 lipids with decreased concentrations in autism samples (Pearson correlation coefficient = 0.96, $P<0.0001$, Figure 4.3).

An analysis of genetic variants linked to each of the three disorders by genetic and genome-wide association studies and collected in the corresponding databases (GRASP, SZDB, SFARI) revealed strong enrichment in genes linked to lipids with decreased concentrations in SZ (hypergeometric test, $P=0.0004$; Figure 4.1 E). This enrichment was robust to the choice of SZ-associated genetic variants from GRASP [Leslie et al., 2014] and SZDB [Wu et al., 2016] databases (Figure 4.1 E). A functional analysis of these genes yielded 19 pathways containing excesses of SZ-linked genetic variants (hypergeometric test, BH-corrected $P<0.01$). These pathways included "phosphatidylinositol signaling", which was also identified in lipid class analysis (Figure 4.1 C), as well as "long term depression" and "glutamatergic synapse".

4.2.2 Membrane fluidity alterations inferred based on lipidome composition

The extent of membrane fluidity, an important biological parameter characterizing membrane biochemical and physiological properties, could be inferred from the membrane lipidome composition [Lauwers et al., 2016, Gennis, 1989, Heimburg, 2007]. We assessed membrane fluidity alterations in ASD, SZ and DS, using the following lipidome features: (i) the total cholesterol concentration, (ii) the cholesterol to phospholipids ratio, the proportion of (iii) saturated phospholipids, (iv) glycerophosphocholines and (v) glycerophosphoethanolamines, which negatively correlating with membrane fluidity, as well as the relative unsaturation state of fatty acyl chains of (vi) GP, (vii) PC and (viii) PE, which positively correlating with membrane fluidity [Lauwers et al., 2016, Gennis, 1989, Heimburg, 2007]. In disorders, SZ samples demonstrated increased membrane fluidity compared to age-matched controls, while ASD and DS samples had decreased membrane fluidity (Figure 4.4).

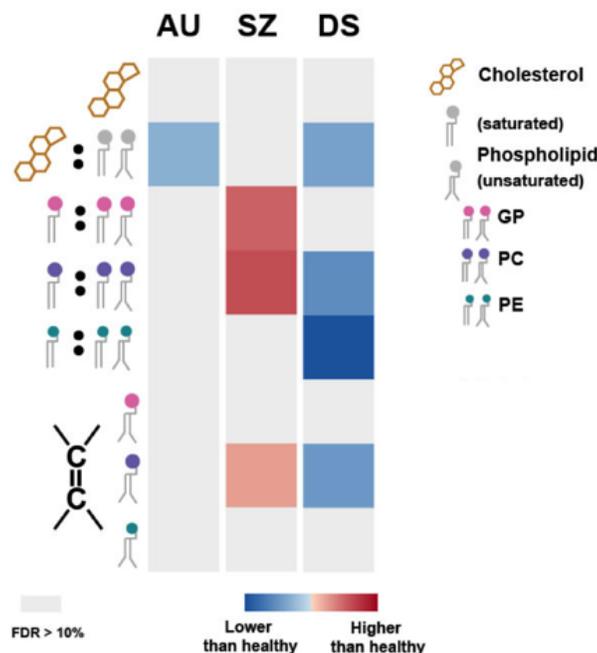


Figure 4.4: Predicted membrane fluidity levels in disorders [Yu et al., 2018]. Relative membrane fluidity estimates in three disorders using the following features (from top to bottom): (i) total cholesterol concentration, (ii) ratio of cholesterol to phospholipid concentration, proportion of saturated (iii) phospholipids, (iv) PC, (v) PE and relative unsaturation degree of fatty acyl of (vi) GP, (vii) PC and (viii) PE, respectively. Each cell shows \log_2 -transformed fold change in disorder compared to age-matched controls. Gray cells represent the absence of significant changes (Wilcoxon rank sum test, BH-corrected $P > 0.1$).

4.3 Discussion

Lipid concentration changes were reported for specific lipid classes in many cognitive disorders, including SZ [Kaddurah-Daouk et al., 2007, McEvoy et al., 2013, Orešič et al., 2012, Solberg et al., 2016, Tessier et al., 2016, Schwarz et al., 2008, Sethi et al., 2017], ASD [El-Ansary et al., 2011, Kim et al., 2010], DS [Adelekan et al., 2012], bipolar disorder [Schwarz et al., 2008, Chung et al., 2007] and Alzheimer’s disease [Chan et al., 2012]. In our study, 27 SZ, 17 ASD and five DS patients samples were measured together with the control group in a random order. Overall, despite limited sample numbers, our results confirm the existence of substantial changes in the PFC lipidome composition in all three disorders. Notably, we demonstrate that SZ-associated genetic variants are robustly enriched in genes linked to lipids showing decreased concentrations in SZ in our study. Thus, at least in SZ, lipid concentration changes might be linked to disease causes.

All three diseases show parallel lipid concentration alterations in two pathways: "retrograde endocannabinoid signaling" and "glycerophospholipid metabolism". Retrograde endocannabinoid signaling, also known as endocannabinoid (eCB) signaling, was shown to play an important role in the control of emotional responses, contextual behavior reactions and social interactions [Karhson et al., 2016, Wei et al., 2017]. Changes in concentrations of endocannabinoids and cannabinoid receptors were reported in SZ [Eggan et al., 2008, Muguruza et al., 2013], ASD [Chakrabarti et al., 2015], and Huntington's disease [Pazos et al., 2008]. In our analysis, lipids with decreased concentrations in autism samples in DS1 and DS2 are both enriched in eCB signaling pathway (Figure 4.3). Of the two well-characterized endocannabinoid compounds involved in eCB, arachidonylethanolamide (AEA) shows a concentration decrease in autism DS2 (Figure 4.3), consistent with results obtained in model organisms [Karhson et al., 2016, Wei et al., 2017]. The second endocannabinoid, 2-arachidonoyl glycerol (2-AG), shows increased concentration in ASD in both DS1 and DS2, as well as in SZ. Consistently, a 2-AG concentration increase was reported in the PFC of SZ patients [Muguruza et al., 2013]. Concentrations of other lipid components of eCB signaling pathway, eCB biosynthesis and degradation products, were not assessed previously. Our results indicate that eCB biosynthesis intermediates, i.e. glycerophosphocholines (PC), glycerophosphoethanolamines (PE) and diacylglycerol (DAG) are mainly decreased in all three disorders, although DAG concentration increases in DS2 - the only contradictory result between datasets among seven detected metabolites. The degradation products, i.e. prostaglandin H2 (PG-G) and arachidonate (AA) are decreased and increased respectively (Figure 4.1 D, Figure 4.3 C).

For the other pathway showing an overall lipid concentration decrease in the three disorders, "glycerophospholipid metabolism" (GP metabolism), lipid concentration decreases and increases were both reported in SZ [Kaddurah-Daouk et al., 2007, Orešič et al., 2012, Tessier et al., 2016, Schwarz et al., 2008, Sethi et al., 2017]. Our results indicate that, similar to eCB signaling, the disruption of GP metabolism might be a common feature of cognitive dysfunctions. Furthermore, GP metabolism includes metabolism of PC, PE, and DG participating in eCB biosynthesis. Decreased levels of PC and PE were reported in SZ [Berger et al., 2006, Schmitt et al., 2004]. Thus, the two pathways showing an overall lipid concentration decrease in the disorders are linked.

Among lipidome changes particular to each disease, some could be linked to the reported changes. For instance, a lower concentration of cardiolipin, the critical component of the inner mitochondrial membrane, in SZ (Figure 4.1 C) matches elevated anti-cardiolipin antibody levels reported in SZ patients [Chang et al., 2011, Careaga et al., 2013]. Similarly, elevated

levels of triacylglycerols (TG) in DS match the reported overexpression of genes involved in energy consumption and oxidative stress in DS patients [[Helguera et al., 2013](#)], as TG represent one of the major energy sources, and were linked to an increase of lipid peroxidation markers [[Pérez-Rodríguez et al., 2015](#)].

Chapter 5

Lipidome evolution in mammalian tissues

In this chapter, we performed large-scale analysis of the lipidome evolution in six tissues of 32 species representing primates, rodents, and bats. While changes in genes' sequence and expression accumulate proportionally to the phylogenetic distances, <2% of the lipidome evolves this way. Yet, lipids constituting these 2% are clustered in specific functions shared among all tissues. Among species, human show the largest amount of species-specific lipidome differences. Many of the uniquely human lipidome features localize in the brain cortex and cluster in specific pathways implicated in cognitive disorders.

5.1 Materials and Methods

5.1.1 Samples

Human samples were obtained from the NICHD Brain and Tissue Bank for Developmental Disorders at the University of Maryland, the Netherlands Brain Bank, and the Chinese Brain Bank Center (CBBC, <http://cbbc.scuec.edu.cn>, Wuhan, China). In accordance with the protocols of these institutions, specific permission for brain autopsies and the use of brain tissue for research purposes was given by the donors or their relatives. The use of human autopsy tissue is considered nonhuman subject research and is IRB exempt under NIH guidelines. All subjects were defined as healthy with respect to the sampled tissue by forensic pathologists at the corresponding tissue bank. All subjects suffered sudden death with no prolonged agony state.

Primate samples were obtained from Simian Laboratory Europe (SILABE) in Strasburg, research unit CNRS-MNHM 717 in Brunoy, France, the German Primate Center (DPZ) in Goettingen, the Max Planck Institute for Anthropology in Leipzig, Germany, the Suzhou Experimental Animal Center in China, the Anthropological Institute and Museum of the University of Zurich-Irchel, Switzerland, and the Biomedical Primate Research Centre in the Netherlands. All nonhuman primates used in this study suffered sudden deaths for reasons other than their participation in this study and without any relation to the tissue used.

Rodent samples were obtained from the University of Rochester Biology Department, MDC Berlin, Department of Zoology and Entomology, University of Pretoria, and the animal center at the Shanghai Institute for Biological Sciences. All mice were from the C57/BL6 strain with no genetic modifications. Bat samples were obtained from Kunming Institute of Zoology of the Chinese Academy of Sciences. The use and care of the animals in this research was reviewed and approved by the Biological Research Ethics Committee, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences. All samples were lawfully acquired and their retention and use were in every case in compliance with national and local laws and regulations. Additionally, all nonhuman samples acquisition, retention and use were in every case in accordance with the Institute for Laboratory Animal Research (ILAR) Guide for the Care and Use of Laboratory Animals.

5.1.2 Normalization of concentrations

To limit potential technical artifacts, mass spectrometric peaks that had zero values in $>50\%$ of individuals were removed. All peak concentration values were then normalized by the internal standard (IS) - 1,2-diheptadecanoyl-sn-glycero-3-phosphocholine concentration levels and \log_2 transformed. Before the \log_2 transformation, we added 1 to the concentration values in order to avoid infinite values. Missing concentration values were ignored in further analyses. Normalized concentrations were transformed into Z-scores for each peak.

5.1.3 Peak annotation

Peaks were matched to the LIPID MAPS database as described [Fahy et al., 2009]. Peaks showing a correlation >0.7 and a difference in RT of <0.05 that matched the same hydrophobic compound in the database search were merged together. This procedure resulted in the merging of 427–880 peaks in different tissues into composite peaks of size of up to six. As a result, the data sets contained between 5,313 and 13,067 compounds in different tissues and

ionization modes and between 13,089 (cerebellum) and 20,669 (heart) compounds in both ionization modes together (Figure 5.2 B).

We used LIPID MAPS [Fahy et al., 2009], HMDB [Wishart et al., 2012], and KEGG [Kanehisa et al., 2017] database annotations to assign each identified compound to its lipid class, subclass, and pathway. Unannotated peaks were excluded from further analysis. About 1,231 peaks were detected and annotated across all six tissues, and were further used to estimate the relationship among samples with a multidimensional scaling algorithm, after quantile normalization of concentration values (Figure 5.2 C).

We tested the overrepresentation of lipid classes, subclasses, and pathways in a given group of peaks (lipids) using a hypergeometric test corrected for multiple testing by the Benjamini and Hochberg method (Figure 5.3 E and 5.4 E).

5.1.4 Identification of lipids and genes that follow phylogeny

To estimate lipid concentration changes during evolution, we fitted a linear model for each lipid using concentration differences between any two samples as dependent variables and phylogenetic distances for the same samples as independent variables. Phylogenetic distances were obtained from the phylogenetic tree <http://www.timetree.org/> [Kumar et al., 2017]. Based on the linear models, for each lipid we estimated a coefficient of determination (R^2), which in this case is a measure of the relationship between concentration differences and phylogenetic distances. Additionally, we quantified the phylogenetic signal in lipid evolution using Blomberg’s K as described in [Ma et al., 2015]. As both approaches yielded similar results we defined phylogeny-dependent lipids as the ones with $R^2 \geq 0.1$ (Figure 5.3 A). To determine the validity of this cutoff, we calculated Spearman correlations between concentration differences and phylogenetic distances. All lipids with concentration differences between species following phylogenetic distances demonstrated positive correlation coefficients. Moreover, we obtained qualitatively similar results at a different R^2 cutoff ($R^2 \geq 0.2$).

To construct the background distribution of R^2 values, we performed 1,000 permutations of species labels across samples within clades, and for all samples as well, and repeated R^2 calculation procedure described above for each permutation (Figure 5.3 A). The same procedure was carried out in Blomberg’s K distribution analysis.

Additionally, we performed permutations for individual lipids. Specifically, in each tissue, we performed 1,000 permutations of species labels, while preserving data structure, that is, all individuals representing one species were switched to a new species label together. We then defined the permutation P value as the proportion of permutations where the coefficient of determination (R^2) or Blomberg’s K was higher than or equal to the original one. All lipids

defined as phylogeny-dependent by either R^2 or Blomberg's K passed the permutation P value cutoff of 0.05, and $>93\%$ passed the permutation P value cutoff of 0.01 (Figure 5.1).

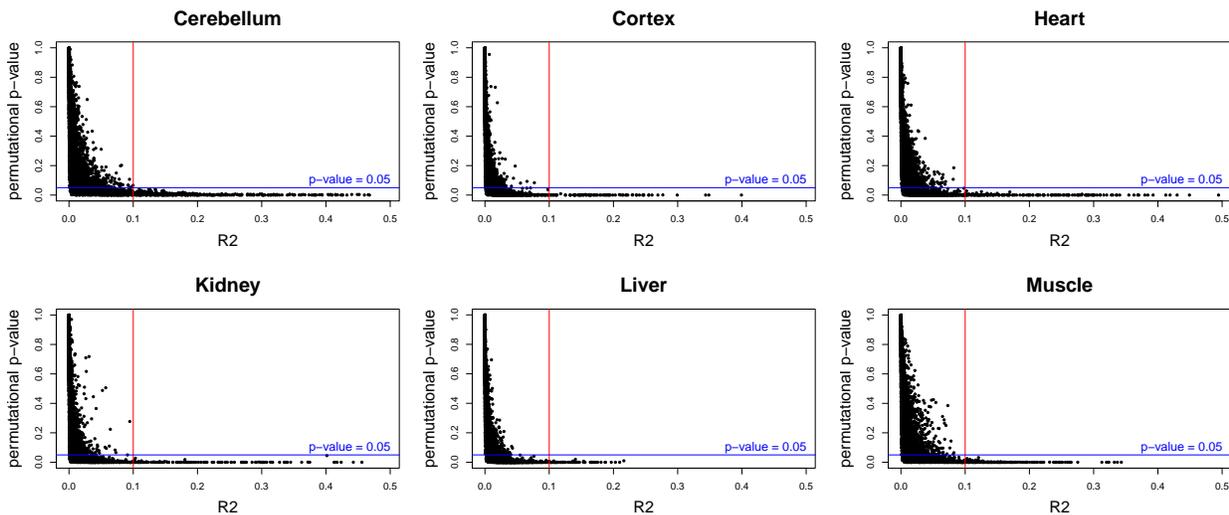


Figure 5.1: Dependence between R^2 values and R^2 permutation p-values calculated for each lipid separately in six tissues [Khrameeva et al., 2018].

The exact same analysis was performed for the gene expression data set GSE43013 [Fushan et al., 2015]. For the data set preprocessing, genome annotations were obtained from Ensembl, release 90. Paired-end reads were downloaded from SRA and mapped using STAR v2.5.3a with default settings. Average gene expression levels were calculated as FPKM using RSEM v.1.2.31. Orthologs were obtained from Ensembl biomaart, release 90. To eliminate unambiguity, genes that had more than one ortholog in the same species were removed. Further, FPKM values were \log_2 transformed, and quantile normalization was applied. Genes that had missing values in $\geq 50\%$ of individuals were removed. To compare percentages of phylogeny-dependent protein-coding genes and lipids, we analyzed transcriptome and lipidome data in a matching subset of three tissues of eight species: *H. sapiens*, *P. troglodytes*, *M. mulatta*, *M. musculus*, *R. norvegicus*, *M. auratus*, *C. porcellus*, *H. glaber* (Figure 5.3 B).

Additionally, we repeated the same analysis for the gene expression data set GSE30352 [Brawand et al., 2011]. To compare percentages of phylogeny-dependent protein-coding genes and lipids, we analyzed transcriptome and lipidome data in a matching subset of four tissues of eight species: *P. troglodytes*, *M. mulatta*, *H. sapiens*, *C. capucinus*, *C. jacchus*, *M. musculus*, *H. glaber*, and *H. armiger* for lipidome data and *P. troglodytes*, *M. mulatta*, *H. sapiens*, *M. musculus*, *G. gorilla*, *P. pygmaeus*, *M. domestica*, and *O. anatinus* for transcriptome data.

5.1.5 Identification of protein coding sequences that follow phylogeny

Orthologous coding sequences for mammal species were obtained from the OrthoMaM database (v9) [Douzery et al., 2014]. We estimated pairwise distances between two orthologous sequences using the Jenson-Shannon divergence measure implemented in Spaced words software [Horwege et al., 2014]. To fit linear models, the obtained Jenson-Shannon divergences were used as dependent variables, and phylogenetic distances between the same species were used as independent variables. Based on the linear models, for each protein sequence we estimated a coefficient of determination (R^2). Further steps were performed exactly as for lipids (see above, Figure 5.6 A).

5.1.6 Analysis of lipid concentration differences within and between species

To construct the control lipid group, we selected a subset of remaining lipids so as to ensure its size and concentration distribution were the same as those of phylogeny-dependent lipids.

To estimate intraspecies differences of lipid concentrations among individuals, we calculated pairwise Spearman correlations of concentrations between samples derived from the same species (Figure 5.3 D). To estimate interspecies differences of lipid concentrations among individuals, we calculated pairwise Spearman correlations of concentrations between samples derived from different species (Figure 5.3 C). Since evolutionarily distant species might yield lower interspecies Spearman's correlation coefficients than the closely related ones, we additionally normalized the obtained values by the average phylogenetic distances between this species and the rest. For all lipids, we defined intraspecies concentration differences as 1 minus the intraspecies Spearman correlation and interspecies concentration differences as 1 minus the interspecies Spearman correlation. All calculations were performed for each tissue separately.

To confirm our observations, we repeated the analysis for phylogeny-dependent lipids defined using Blomberg's K as described in [Ma et al., 2015] at $K \geq 1.2$ cutoff.

5.1.7 Identification of species-specific lipids

For each lipid, we compared concentrations in one species against all other species using the Wilcoxon test. If the Wilcoxon test P value was lower than the chosen cutoff (0.01), the lipid was defined as species-specific (Figure 5.4 A). To balance the number of samples

per species, we randomly selected three samples in each species, repeated the procedure 100 times, and calculated the sum number of species-specific lipids among 100 iterations for each species. Because the analysis was applied to each species and each tissue separately, it was limited to 23 out of 32 lineages containing at least three biological replicates in a given tissue. Since evolutionarily distant species might yield higher numbers of species-specific lipids than the closely related ones, we normalized the obtained numbers of species-specific lipids by the average phylogenetic distances between this species and the rest. To test whether the observed number of species-specific lipids was higher than expected by chance, we permuted species labels between samples 100 times, each time calculating the number of species-specific lipids as described earlier. For each species, we then calculated the permutation P value as the proportion of permutations resulting in an equal or greater number of significant lipids at the chosen nominal significance cutoff (0.01, Figure 5.4 B). The analysis was performed for lipids measured in the negative ionization mode only, which included the majority of detected lipids.

To confirm our observations with an alternative approach, we estimated the number of species-specific lipids using the Ornstein-Uhlenbeck (OU) model, as implemented in R package `ouch` [Butler and King, 2004]. To find whether changes in optimal lipid concentration have occurred at a particular branch of the phylogenetic tree, we tested the null hypothesis that all branches share the same optimum parameter against the alternative hypothesis that there is a distinct optimum at a different phylogenetic branch corresponding to a particular species. For each lipid, we built an S_0 model (one global optimum across all species) and S_1, S_2, \dots, S_x models (different optimums for particular species x). Based on these models, we used the likelihood ratio test between a species-specific selective regime S_x model and the S_0 model. The likelihood ratio between S_x and S_0 had an asymptotic χ^2 distribution (with one degree of freedom) from which a P value could be calculated. If the P value was lower than the chosen cutoff (0.01), the lipid was defined as species-specific. To balance the number of samples per species, we randomly selected three samples in each species, repeated the procedure 100 times, and calculated the sum number of species-specific lipids among 100 iterations for each species. To test whether the observed number of species-specific lipids was higher than expected by chance, we performed permutations exactly as described earlier for species-specific lipids defined using the Wilcoxon test.

5.1.8 Comparison with published data set

We defined lipids with human-specific concentration changes as lipids demonstrating Wilcoxon test P values (for human-vs-other-species comparisons) of < 0.05 in 95% of boot-

straps in one tissue. Bootstraps were performed via the random selection of three individuals per species, repeated 100 times. Lipids with missing values in $\geq 50\%$ of permutations were removed. To compare the remaining lipids with a published data set [Bozek et al., 2015] we selected lipids with a common annotation in both data sets. In cases when multiple LIPIDMAPS IDs were assigned to each lipid, we selected lipids that shared the greatest number of LIPIDMAPS IDs in the two data sets. We plotted concentration distributions for these lipids (Figure 5.4 D). Additionally, we calculated \log_2 -fold changes between human lipid concentrations and average lipid concentrations of chimpanzees, macaques, and mice (Figure 5.4 C). To check if \log_2 -fold change values were in agreement between data sets, we performed the Fisher’s test.

5.1.9 Human-specific gene expression changes

We defined lipids with human-specific concentration changes as lipids demonstrating Wilcoxon test P values (for human-vs-other-species comparisons) of < 0.01 in 75% of bootstraps (three random individuals per species) in one tissue. We selected enzymes directly linked to the lipids according to the KEGG database [Kanehisa et al., 2017], and defined protein-coding genes with human-specific expression as those showing Wilcoxon test P values of < 0.05 for humans vs. other primate comparisons in one tissue, according to gene expression values in the public data set GSE30352 [Brawand et al., 2011] (Figure 5.5 A).

We tested the overrepresentation of KEGG pathways in each gene group using a hypergeometric test corrected for multiple testing by the Benjamini and Hochberg method (Figure 5.5 B and C).

5.2 Results

5.2.1 Lipidome data description

We analyzed lipid concentrations in the liver (LV), muscle (ML), kidney (KD), heart (HT), brain cortex (CX), and cerebellum (CB) using 669 samples obtained from 32 mammalian species representing three phylogenetic clades: rodents, primates, and bats [Bozek et al., 2017] (Figure 5.2 A). Our analysis was based on the quantification of $> 18,000$ (median = 18,708) mass spectrometry features in each tissue corresponding to hydrophobic compounds (lipids) detected in at least 50% of the individuals of each species. Among these lipids, close to 7,000 (median = 6,994) were annotated in each tissue using computational matching to the hydrophobic compound database [Fahy et al., 2009] (Figure 5.2 B). A multidimensional

scaling analysis revealed that the samples segregated predominantly according to organ origin (Figure 5.2 C).

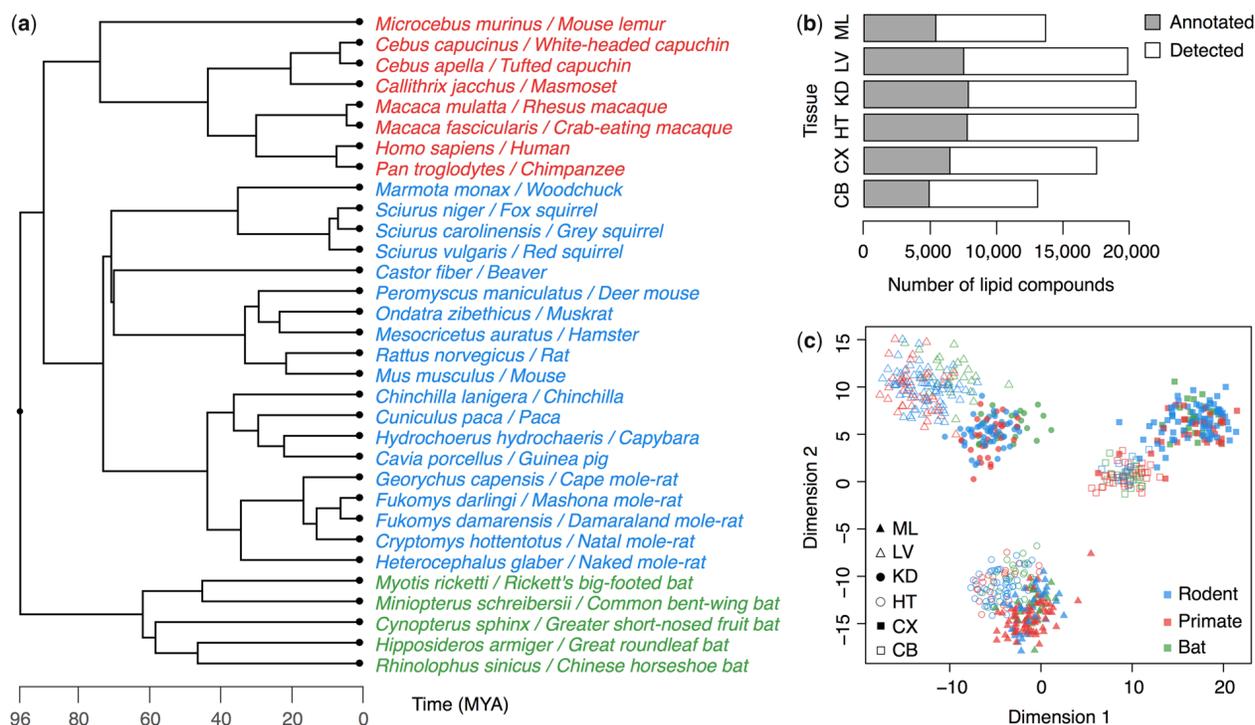


Figure 5.2: Lipidome data overview [Khrameeva et al., 2018]. (a) Phylogenetic tree of 32 mammalian species used in this study. The colors indicate three main represented clades: rodents (blue), primates (red), and bats (green). (b) Numbers of lipids detected in each tissue. Gray bars represent annotated lipids. Here and later the tissues are labeled by a two-letter code: CB, cerebellum; CX, cortex; HT, heart; KD, kidney; LV, liver; ML, muscle. (c) The relationship among samples are based on concentrations of 1,231 annotated lipids plotted in two dimensions using a multidimensional scaling algorithm. Symbols represent tissues, colors represent clades, and points represent individual samples.

5.2.2 Lipidome differences and phylogenetic distances among species

Most of the genome sequence differences and a substantial proportion of gene expression differences among species tend to scale linearly with phylogenetic distances. This indicates that the majority of these changes might have no effect on survival and reproductive abilities [Lanfear et al., 2010, Kuraku et al., 2016]. We used linear regression and Blomberg’s K [Blomberg et al., 2003] approaches, analogous to the ones applied in the genome sequence and gene expression analyses, to identify lipids that change linearly with phylogenetic distances

between species (phylogeny-dependent lipids). Surprisingly, on average, only 2% of detected lipids showed this type of concentration differences (R^2 : median = 1.9%, maximum = 9.2% in CB; Blomberg's K: median = 2.1%, maximum = 9.1% in CB; Figure 5.3 A and 5.1).

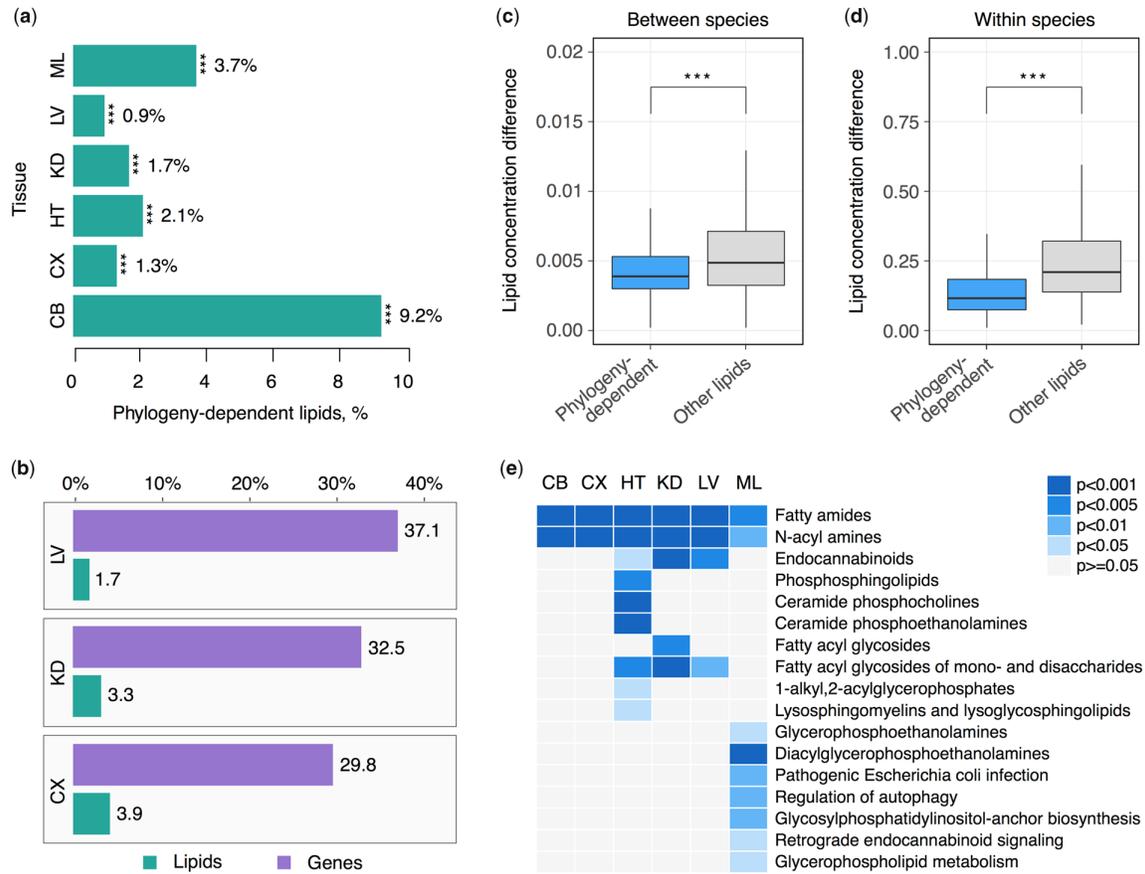


Figure 5.3: Relationship between lipid concentrations and phylogenetic distances [Khrameeva et al., 2018]. (a) Percentages of lipids with concentration differences among species scaling with phylogenetic distances (phylogeny-dependent lipids) in each of six tissues. Stars indicate the significance of the difference between observed number distributions and random expectation (permutations of species labels, $P < 0.001$). (b) Percentages of phylogenetic genes (violet) and lipids (green) defined based on the same criteria in three tissues in a matching subset of eight species. The expression data was taken from [Fushan et al., 2015]. (c) Distribution of concentration differences measured in all pairwise comparisons between species for phylogeny-dependent lipids (blue) and the remaining lipids (light gray). Stars indicate the significance of the difference between the two distributions (two-sided t-test $P < 0.0001$, $n = 66,176$). (d) Distribution of concentration differences among individuals within species for phylogeny-dependent lipids (blue) and the remaining lipids (light gray). Stars indicate the significance of the difference between the two distributions (two-sided t-test $P < 0.0001$, $n = 3,916$). (e) Enrichment of phylogeny-dependent lipids in specific lipid classes and subclasses. Colors indicate BH-corrected enrichment P values.

Notably, the observed proportions of phylogeny-dependent lipids were still significantly greater than expected by chance in all six tissues (permutations, $P < 0.001$). The result was not caused by the selection of species used in the study: a parallel analysis of the transcriptome and lipidome divergence in a subset of three tissues of eight species demonstrated that 30-37% of detected protein-coding transcripts showed phylogeny-dependent expression. For the lipids, however, the median proportion of phylogeny-dependent differences remained low (3.3%) (Figure 5.3 B).

On an average, phylogeny-dependent lipids showed fewer lipid concentration differences among species than the remaining lipids in all six tissues combined (two-sided t-test, $n = 66,176$, $P < 0.0001$, Figure 5.3 C), as well as in each tissue separately, except the cerebellum. Furthermore, phylogeny-dependent lipids showed reduced intraspecies variation (two-sided t-test, $n = 3,916$, $P < 0.0001$, Figure 5.3 D). These results were not caused by the difference in concentration levels between phylogeny-dependent and the remaining lipids or by outlier effects.

Remarkably, despite the relatively low numbers of phylogeny-dependent lipids annotated in each tissue ($N_{\min}=71$, $N_{\text{median}}=148$), they showed unusually strong enrichment in the same lipid class, fatty amides and its subclass N-acyl amides, in all six tissues (hypergeometric test, $n = 3,134$, $3,979$, $5,858$, $5,977$, $5,685$, and $3,384$ in CX, CB, HT, KD, LV, and ML, respectively, $FDR < 0.01$ in each tissue) (Figure 5.3 E). Several other lipid classes, including phosphosphingolipids, and fatty acyl glycosides, were also enriched in phylogeny dependent lipids in one or several tissues.

5.2.3 Species-specific lipidome differences

To further assess the evolutionary dynamics of lipid concentration levels, we identified species-specific differences defined as significant lipid concentration differences between a given species and the others (Figure 5.4 A). The analysis was based on lipid concentration measurements from three individuals per species, sampled randomly among all measured species' individuals. Among all 104 tissue and lineage combinations represented by at least three biological replicates, only three combinations showed significantly more species-specific lipid concentration differences than expected by chance (permutations, $P < 0.05$ at Wilcoxon test $P < 0.01$ cutoff) and five showed a marginally significant trend (permutations, $P < 0.1$ at Wilcoxon test $P < 0.01$ cutoff) (Figure 5.4 B). Of these eight, three were located in the human lineage: two significant ones in the brain cortex and kidney and one marginally significant one in the heart. The use of an alternative approach, the Ornstein-Uhlenbeck (OU) model,

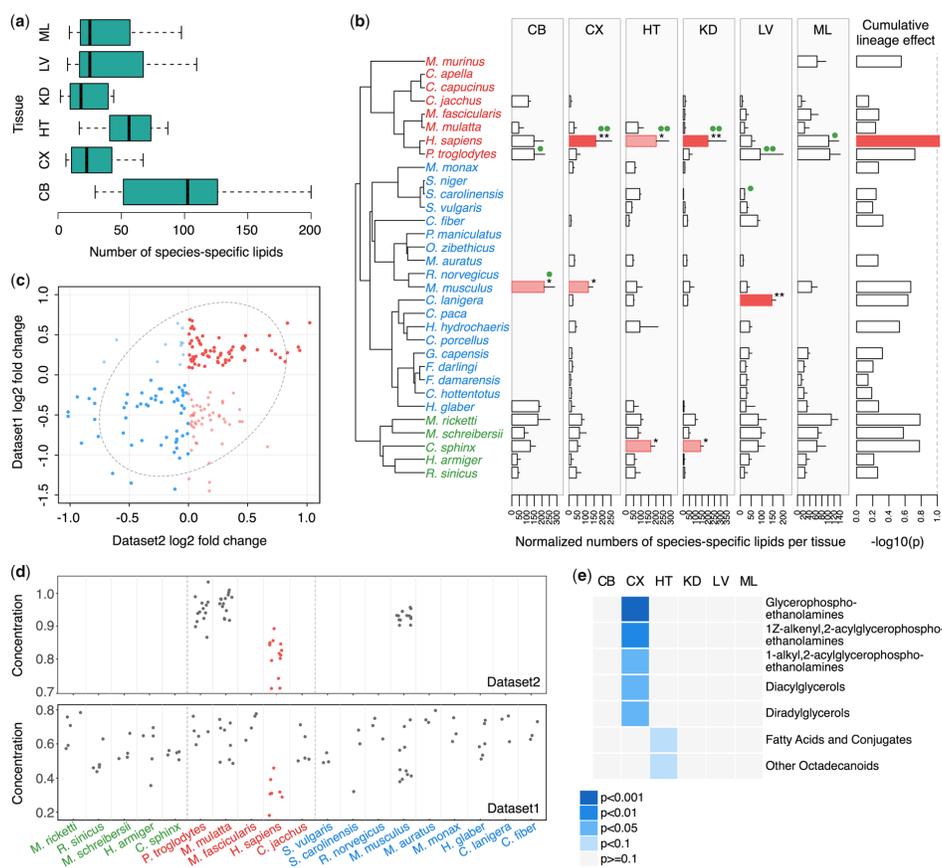


Figure 5.4: Species-specific lipid concentration differences [Khrameeva et al., 2018]. (a) Numbers of lipids showing significant species-specific concentration differences. The distributions show the numbers of such lipids in each of 23 lineages represented by at least three biological replicates in a given tissue. (b) Numbers of lipids showing significant species-specific concentration differences in each of 104 tissue and lineage combinations (normalized by the phylogenetic distances). Error bars show variations of estimates calculated by way of the random sampling of three individuals per species. Stars and bar colors indicate the significance of the difference between observed number distributions and random expectation (permutations, ** and red - $P < 0.05$, * and pink - $P < 0.1$). Green circles indicate the significance of the difference between observed number distributions and random expectation according to the OU model (permutations, oo - $P < 0.05$, o - $P < 0.1$). The rightmost column shows the cumulative lineage effect calculated as an average $-\log_{10} P$ value of the difference between observed and chance numbers of species-specific lipids across tissues. (c) Lipid concentration differences between humans and the other three species (chimpanzee, macaque, and mouse) calculated as \log_2 -transformed fold changes of the average values for 183 lipids showing a significant human-specific concentration difference in our data (Data Set 1) and detected in the published data set (Data Set 2) [Bozek et al., 2015]. Colors indicate signs of \log_2 -transformed fold changes in both data sets. The ellipse shows a 90% confidence interval. (d) An example of the concentrations in the kidney of one lipid (monogalactosyldiacylglycerol, LMGL05010014) shown in the panel (c). (e) Enrichment of lipids with human-specific concentration differences in specific lipid classes and subclasses. Colors indicate BH-corrected enrichment P values.

confirmed three out of three combinations located in the human lineage (permutations, $P < 0.05$ at the OU model $P < 0.01$ threshold). Additionally, the human kidney showed a marginally significant excess of lineage-specific differences (permutations, $P < 0.1$ at the OU model $P < 0.01$ threshold). By contrast, an excess of species-specific lipidome differences in the other lineages was not confirmed by the OU model, except the mouse cerebellum (Figure 5.4 B).

This result indicates the outstanding character of the lipidome evolution in the human lineage, while a robust detection of more subtle species-specific signals in the other lineages might require greater sample sizes. Consequently, the human evolutionary lineage stood out as having the greatest average lineage-specific lipid concentration divergence compared with the other examined mammalian lineages (Figure 5.4 B).

The excess of lipid concentration changes in the human lineage was not caused by differences in the postmortem delay among samples, as estimated based on postmortem delay effects identified in [Bozek et al., 2015]. This excess was not caused by environmental effects either, as human-specific lipid concentration changes were enriched in the brain, yet most lipids composing brain tissue are synthesized in the brain and, therefore, are shielded from environmental and dietary changes by the blood-brain barrier [Sherman and Brophy, 2005, Piomelli et al., 2007].

Additionally, environmental exposure experiments (stress, exercise, and diet factors) conducted in macaques showed that each of the environmental perturbations induced substantially fewer lipid concentration changes compared with the lipidome differences observed between chimpanzees and humans [Bozek et al., 2015]. Consistently, the largest effect of environmental factors was observed in the kidney, and the smallest — in the brain prefrontal cortex [Bozek et al., 2015]. Moreover, human-specific differences detected in this study were in agreement with the differences calculated using a published lipidome data set [Bozek et al., 2015] (Figure 5.4 C and D, Fisher’s test, $OR = 6.5$, $P < 0.0001$).

Lipids showing human-specific concentration changes (HS-lipids) in the brain cortex were clustered in two specific lipid classes, glycerophosphoethanolamines and diradylglycerols and their three subclasses (Figure 5.4 E). Similarly, enzymes linked to HS-lipids demonstrated an excess of human-specific expression changes in the cortex (two-sided t-test, $n = 257$, $P = 0.0004$) (Figure 5.5 A), and were significantly overrepresented in 11 KEGG pathways (the Kyoto Encyclopedia of Genes and Genomes; [Kanehisa et al., 2017]) (hypergeometric test, $n = 511$, BH-corrected $P < 0.0001$, Figure 5.5 B). Of them, three interlinked pathways, glycerophospholipid metabolism, glycerolipid metabolism, and linoleic acid metabolism path-

ways, showed the strongest overrepresentation of both protein-coding genes and lipids showing human-specific expression and concentration levels (Figure 5.5 C).

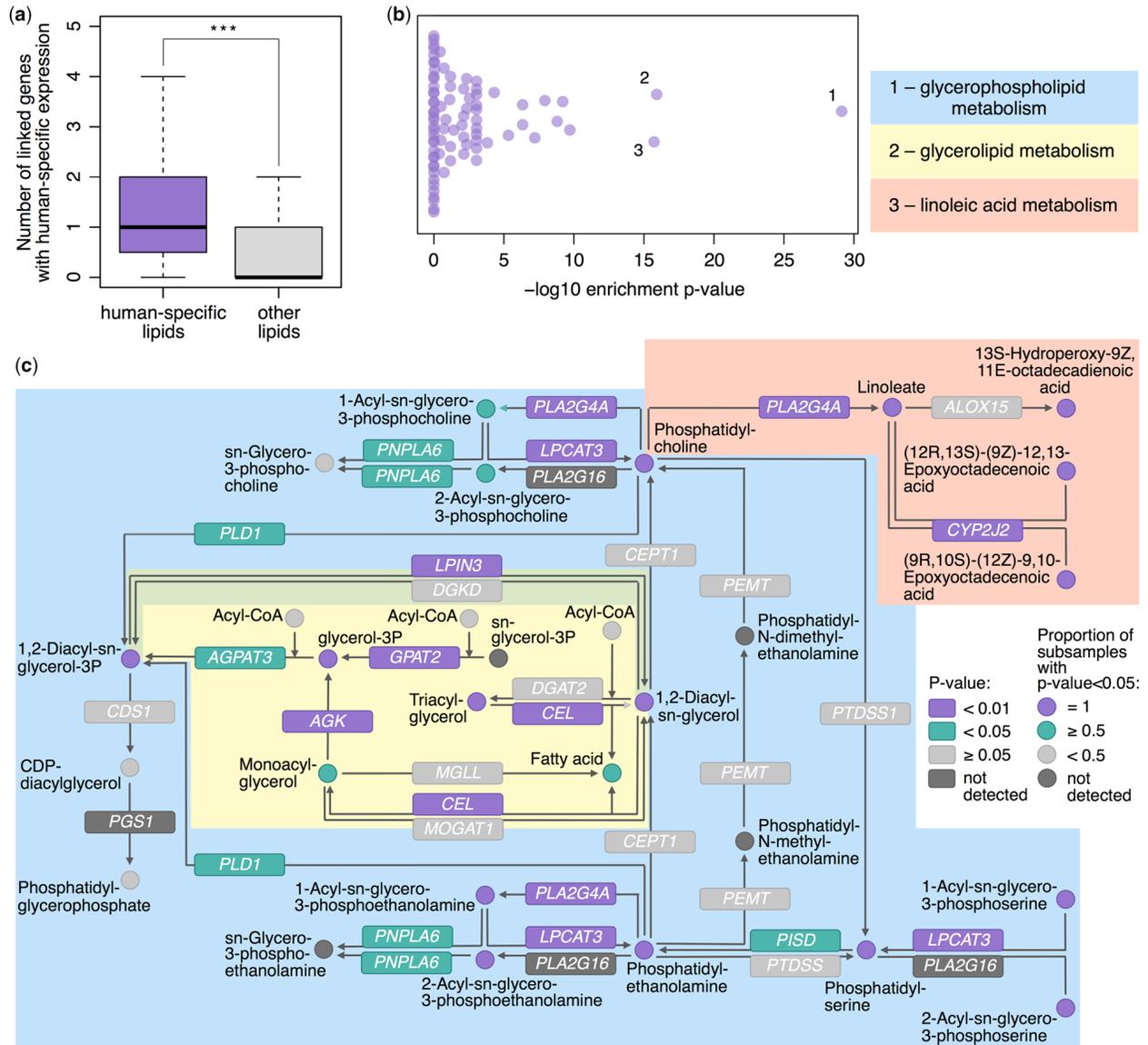


Figure 5.5: Characterization of lipids showing human-specific concentration levels (HS-lipids) in cortex [Khrameeva et al., 2018]. (a) Distribution showing numbers of protein-coding genes with human-specific expression in cortex linked to HS-lipids (violet) and control lipids (gray). Stars indicate the significance of the difference between the two distributions (two-tailed t-test $P < 0.001$, $n = 257$). (b) Enrichment of protein-coding genes linked to HS-lipids in KEGG pathways. Symbols represent pathways. The numbers next to the symbols and the legend above the panel show the top three enriched pathways. (c) The simplified schematic representation of the top three enriched KEGG pathways showing HS-lipids and their linked genes. The background colors indicate the pathways, as in panel (b) legend.

5.3 Discussion

Recent technical advances in measuring lipidome composition in multiple species allow us to assess the evolution at a new level of molecular phenotype: the level of lipid concentrations. Our analysis based on lipidome measurements conducted in six tissues of 32 mammalian species resulted in several remarkable observations. First, in contrast to genetic and gene expression data, where 84.8% and 25.6% of differences among species were proportional to phylogenetic distances, only 1.9% of all lipids' concentrations scaled with the phylogeny (Figure 5.6 A). Moreover, unlike genetic and gene expression changes, lipid concentration changes that followed the phylogeny did not show characteristic features of a neutral evolutionary model. Specifically, while much more abundant genetic and gene expression changes did not

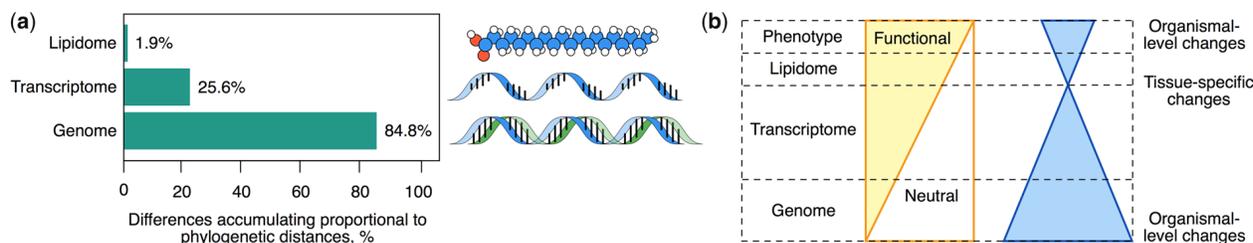


Figure 5.6: Evolution at lipidome, transcriptome and genome levels [Khrameeva et al., 2018]. (a) Percentages of lipids, protein-coding transcripts and genes with differences among species accumulating proportionally to phylogenetic distances. (b) Schematic representation of the suggested lipidome evolution properties in comparison to the other levels of molecular and organismal phenotype. Left: relative proportions of neutral and functional differences among species. Right: proportion of evolutionary differences shared among tissues. The vertical axis shows the relevant levels of organismal organization. The inversion of this proportion between the transcriptome and lipidome levels reflects the notion that ubiquitously present transcriptome differences are mainly neutral and lipidome ones are mainly functional.

cluster in any functional categories (hypergeometric test, $n = 5,206, 5,219, 5,120, 5,206,$ and $5,062$ in CX, CB, HT, KD, and LV, respectively, BH-corrected $P > 0.1$), the few lipids that followed the phylogeny were grouped in the same particular lipid class, fatty amides, in all six tissues. These lipids showed smaller amplitude of concentration differences among species and varied less within species than the rest of the lipidome. Taken together these two observations, the remarkable agreement between tissues, which otherwise have very different lipidome compositions, and reduced concentration variation, suggests the importance of fatty amides in as yet unexplained functions related to an evolutionary “clock-like” lipidome divergence in all tissues. The review of the known features of fatty amides revealed that they are indeed found in all tissues of mammal species [Zoerner et al., 2011]. However, the evo-

lution of fatty amides and, in particular, of their most abundant subclass endocannabinoids, represented by 15% of all fatty amides detected in six tissues, has been studied at the genetic level only [McPartland et al., 2006, Elphick, 2012]. Like the pharmacologically active compounds in marijuana or cannabis, endocannabinoids exert their effects by binding to and activating specific cannabinoid receptors CB1 and CB2. The endocannabinoids produce neurobehavioral effects and have key neurotransmitter roles in the central nervous system, especially in the perception of pain, stress, and anxiety, in energy balance and in appetite control [Piomelli, 2003]. Moreover, endocannabinoids have anti-inflammatory and anticancer properties [De Petrocellis et al., 2000].

We further assessed the possible relationship between the lipidome and species' phenotypes by identifying lipid concentration changes unique to each species. Even though the number of individuals per species was limited, we were able to examine 23 lineages represented by at least three biological replicates per tissue, and detected a significant excess of lipidome changes in the human, mouse, and chinchilla lineages, as well as a marginally significant trend in the two bat lineages. Notably, the human lineage stood out among the rest by showing the greatest cumulative species-specific divergence. Due to the small numbers of individuals examined in our study, it is likely that many species-specific lipid concentration changes remained undetected. Still, it is noteworthy that among 23 lineages used in the analysis, the human one showed the most pronounced excess of lipid concentration changes. Furthermore, of all species, only humans showed a significant excess of lipid concentration changes in the brain cortex. This result agrees well with observations of accelerated lipidome evolution in the human brain [Bozek et al., 2015, Li et al., 2017].

Human-specific lipidome changes were most pronounced in the brain cortex and kidney, but only cortical changes clustered in distinct functional pathways, particularly in glycerolipid, glycerophospholipid, and linoleic acid metabolism. These pathways were implicated in a number of neurodegenerative disorders such as Alzheimer's disease [Snowden et al., 2017], Parkinson's disease [Cheng et al., 2011], and neurodegeneration with brain iron accumulation [Morgan et al., 2006], as well as other nervous system disorders: hereditary spastic paraplegia, congenital myasthenic syndrome, Fabry disease, pyridoxine-dependent epilepsy, and Sjögren-Larsson syndrome [Kanehisa et al., 2017].

To conclude, our observations indicate that lipid concentrations evolve differently compared with genome sequences and gene expression levels. We speculate that lipid evolution represents phenotypic differences between species more closely than genomic and gene expression differences do, with a greater proportion of differences accumulating among species

representing functional changes shared among tissues (Figure 5.6 B), while tissue-specific lipidome differences might represent species-specific adaptations.

Chapter 6

Conclusions

In the absence of common genetic determinants of neurocognitive disorders, such as ASD and SZ, concentrations of low weight molecular compounds in urine, blood plasma, and brain might provide a more direct link to the molecular signature of the disease. Lipids and metabolites are essential structural components and signaling molecules of the human brain, yet our understanding of their involvement in brain functions and dysfunctions is only beginning to emerge.

In my thesis, I show that despite a relatively small sample size of our study, limited to 32 autism patients and 40 age-matched control individuals, 15% of detected metabolites showed significant concentration differences between autism and control samples in the prefrontal cortex grey matter. Those metabolic alterations cluster in specific pathways, suggesting non-random and possibly functionally important character of the disease-related changes. Additionally, we found that based on concentrations of 200 metabolites, autism samples can be separated from the controls with more than 95% accuracy using machine learning algorithms. These results indicate that brain metabolism is substantially altered in ASD. Remarkably, many of these alterations, both at the pathway level and at the level of individual metabolites, coincide with the differences reported in urine and blood of ASD patients. This fact, together with the high accuracy of autism identification using metabolite concentrations in the brain, underscores the potential of metabolic studies conducted in non-nervous tissues to identify informative diagnostic markers of the disease.

The results of our work open an opportunity for the design of ASD diagnostic tools based on the concentration measurements for a limited set of informative metabolite and lipid compounds in blood or urine samples. Importantly, the concentration changes detected in blood might be informative of the metabolic changes taking place in the brain, as we demonstrated for changes in glutathione and purine metabolic pathways. Creation

of reliable test for ASD diagnostics would ensure timely patients' medical treatment and selection of better-customized treatment routines. Furthermore, it would open the door for widespread metabolic biomarker-based clinical practices, providing the collection of standardized metabolic data from a large number of individuals. This, in turn, would allow the construction of complex predictive models, leading to improved specificity in disease classification and treatment selection. Finally, better knowledge of the molecular changes taking place in the disease will lead to the more rapid development of better medical treatments.

It should be mentioned, however, that design of such diagnostic assays would require extensive validated using independent sets of samples, and preferably include data for the brain, blood and urine samples collected using the same experimental, analytical, and statistical frameworks. Also, the number of samples included in such validation studies should be large, as ASD patients represent a heterogeneous group.

We further considered autism-related changes from the evolutionary perspective, by comparing them to the metabolic differences distinguishing humans from closely related non-human primates: chimpanzees and macaques. We find that some of the pathways affected in autism demonstrated as high as an 8-fold excess of human-specific metabolic differences compared to chimpanzee-specific ones, indicating an accelerated evolution of these pathways on the human lineage. This is potentially interesting, given that autism tends to affect cognitive abilities particularly pronounced in humans and distinguishing our species even from the closely related primates. Thus, the excess of human-specific metabolite concentration differences detected in several of pathways altered in ASD might potentially indicate their involvement in the evolution of cognitive traits unique to humans.

In addition to metabolites, we investigated concentration changes of lipids - non-polar metabolic molecules - in common cognitive disorders. Specifically, lipidome profiling conducted in the prefrontal cortex of cognitively healthy humans, and ASD, schizophrenia and Down syndrome patients, identified multiple lipid concentration changes in the prefrontal cortex in these neurocognitive disorders: 2.6%, 10.4%, and 10.8% of the detected lipids, respectively.

In the absence of the mature annotation of lipid functionality, analogous to gene-based annotation, interpretation of these changes remains a challenge. From the methodological point, to obtain precise annotation of detected metabolite and lipid compounds, we need to perform MS/MS analysis for all informative compounds. This is particularly important in the lipidome studies, as based on one-dimensional MS measurements we cannot assign an annotation to detected compounds unambiguously, given that many combinations of fatty acids with a head group will yield the same molecular mass. Using MS/MS technique it is

possible to resolve the actual fatty composition by analyzing fragmentation products. For example, the combination of fatty acids C18:1/C18:1 or C20:2/C16:0 producing the same precursor ion mass could be successfully resolved after the fragmentation. This is particularly important, as lipids from one lipid class could play quite different functional roles.

Nonetheless, functional annotation of the lipid concentration changes detected in our study in SZ, ASD, and Down Syndrome, even given its limitations, yielded a number of functional pathways particularly affected in these disorders. Some of these pathways, such as the disruption of endocannabinoid signaling in SZ, were reported previously. Furthermore, lipid concentration decrease in SZ could be connected to previously reported genome variants associated with the disease. In the case of ASD, the lipidome differences were supported by an independent dataset, consistently demonstrating an alteration in pathways such as eCB signaling pathway and glycerophospholipid metabolism. These observations indicate that even limited annotation provided by one-dimensional MS data is useful in identification pathways involved in disruption of brain functions and disease development in common cognitive disorders.

Notably, the evolutionary analysis of lipid concentrations altered in diseases indicated that the glycerophospholipid metabolism affected in ASD, schizophrenia, and Down syndrome showed the significant excess of human-specific lipidome differences. This finding further supports the notion that human brain evolution included metabolic changes affecting both polar and non-polar compounds. Notably, the connection between human-specific differences and disease-related ones was reported in ASD at the level of gene expression [Liu et al., 2016] and in schizophrenia at the metabolome level [Khaitovich et al., 2008]. Altogether, those results align well with the hypothesis postulating disruption of recently evolved cognitive mechanisms in schizophrenia and ASD.

Bibliography

- [Adelekan et al., 2012] Adelekan, T., Magge, S., Shults, J., Stallings, V., and Stettler, N. (2012). Lipid Profiles of Children With Down Syndrome Compared With Their Siblings. *PEDIATRICS*, 129(6):e1382–e1387.
- [Adibhatla et al., 2006] Adibhatla, R., Hatcher, J., and Dempsey, R. (2006). Lipids and Lipidomics in Brain Injury and Diseases. *The AAPS Journal*, 08(02):E314.
- [Anney et al., 2010] Anney, R., Klei, L., Pinto, D., Regan, R., Conroy, J., Magalhaes, T. R., Correia, C., Abrahams, B. S., Sykes, N., Pagnamenta, A. T., Almeida, J., Bacchelli, E., Bailey, A. J., Baird, G., Battaglia, A., Berney, T., Bolshakova, N., Bolte, S., Bolton, P. F., Bourgeron, T., Brennan, S., Brian, J., Carson, A. R., Casallo, G., Casey, J., Chu, S. H., Cochrane, L., Corsello, C., Crawford, E. L., Crossett, A., Dawson, G., de Jonge, M., Delorme, R., Drmic, I., Duketis, E., Duque, F., Estes, A., Farrar, P., Fernandez, B. A., Folstein, S. E., Fombonne, E., Freitag, C. M., Gilbert, J., Gillberg, C., Glessner, J. T., Goldberg, J., Green, J., Guter, S. J., Hakonarson, H., Heron, E. A., Hill, M., Holt, R., Howe, J. L., Hughes, G., Hus, V., Iglizzi, R., Kim, C., Klauck, S. M., Kolevzon, A., Korvatska, O., Kustanovich, V., Lajonchere, C. M., Lamb, J. A., Laskawiec, M., Leboyer, M., Le Couteur, A., Leventhal, B. L., Lionel, A. C., Liu, X.-Q., Lord, C., Lotspeich, L., Lund, S. C., Maestrini, E., Mahoney, W., Mantoulan, C., Marshall, C. R., McConachie, H., McDougle, C. J., McGrath, J., McMahon, W. M., Melhem, N. M., Merikangas, A., Migita, O., Minshew, N. J., Mirza, G. K., Munson, J., Nelson, S. F., Noakes, C., Noor, A., Nygren, G., Oliveira, G., Papanikolaou, K., Parr, J. R., Parrini, B., Paton, T., Pickles, A., Piven, J., Posey, D. J., Poustka, A., Poustka, F., Prasad, A., Ragoussis, J., Renshaw, K., Rickaby, J., Roberts, W., Roeder, K., Roge, B., Rutter, M. L., Bierut, L. J., Rice, J. P., Salt, J., Sansom, K., Sato, D., Segurado, R., Senman, L., Shah, N., Sheffield, V. C., Soorya, L., Sousa, I., Stoppioni, V., Strawbridge, C., Tancredi, R., Tansey, K., Thiruvahindrapduram, B., Thompson, A. P., Thomson, S., Tryfon, A., Tsiantis, J., Van Engeland, H., Vincent, J. B., Volkmar, F., Wallace, S., Wang, K., Wang, Z., Wassink, T. H., Wing, K., Wittmeyer, K., Wood, S., Yaspan, B. L., Zurawiecki, D., Zwaigenbaum, L., Betancur, C., Buxbaum, J. D., Cantor, R. M., Cook, E. H., Coon, H., Cuccaro, M. L., Gallagher, L., Geschwind, D. H., Gill, M., Haines, J. L., Miller, J., Monaco, A. P., Nurnberger, J. I., Paterson, A. D., Pericak-Vance, M. A., Schellenberg, G. D., Scherer, S. W., Sutcliffe, J. S., Szatmari, P., Vicente, A. M., Vieland, V. J., Wijsman, E. M., Devlin, B., Ennis, S., and Hallmayer, J. (2010). A genome-wide scan for common alleles affecting risk for autism. *Human Molecular Genetics*, 19(20):4072–4082.
- [Antonny, 2011] Antonny, B. (2011). Mechanisms of Membrane Curvature Sensing. *Annual Review of Biochemistry*, 80(1):101–123.

- [Araki et al., 2009] Araki, K., Takino, T., Ida, S., and Kuriyama, K. (2009). Alteration of amino acids in cerebrospinal fluid from patients with Parkinson’s disease and spinocerebellar degeneration. *Acta Neurologica Scandinavica*, 73(2):105–110.
- [Ariga et al., 2008] Ariga, T., McDonald, M. P., and Yu, R. K. (2008). *Thematic Review Series: Sphingolipids*. Role of ganglioside metabolism in the pathogenesis of Alzheimer’s disease—a review,. *Journal of Lipid Research*, 49(6):1157–1175.
- [Bankir and Yang, 2012] Bankir, L. and Yang, B. (2012). New insights into urea and glucose handling by the kidney, and the urine concentrating mechanism. *Kidney International*, 81(12):1179–1198.
- [Bell et al., 2004] Bell, J., MacKinlay, E., Dick, J., MacDonald, D., Boyle, R., and Glen, A. (2004). Essential fatty acids and phospholipase A2 in autistic spectrum disorders. *Prostaglandins, Leukotrienes and Essential Fatty Acids*, 71(4):201–204.
- [Berger et al., 2006] Berger, G. E., Smesny, S., and Amminger, G. P. (2006). Bioactive lipids in schizophrenia. *International Review of Psychiatry*, 18(2):85–98.
- [Bitar et al., 2018] Bitar, T., Mavel, S., Emond, P., Nadal-Desbarats, L., Lefèvre, A., Mattar, H., Soufia, M., Blasco, H., Vourc’h, P., Hleihel, W., and Andres, C. R. (2018). Identification of metabolic pathway disturbances using multimodal metabolomics in autistic disorders in a Middle Eastern population. *Journal of Pharmaceutical and Biomedical Analysis*, 152:57–65.
- [Blomberg et al., 2003] Blomberg, S. P., Garland, T., and Ives, A. R. (2003). TESTING FOR PHYLOGENETIC SIGNAL IN COMPARATIVE DATA: BEHAVIORAL TRAITS ARE MORE LABILE. *Evolution*, 57(4):717–745.
- [Boyle et al., 2017] Boyle, E. A., Li, Y. L., and Pritchard, J. K. (2017). An Expanded View of Complex Traits: From Polygenic to Omnigenic. *Cell*, 169(7):1177–1186.
- [Bozek et al., 2017] Bozek, K., Khrameeva, E. E., Reznick, J., Omerbašić, D., Bennett, N. C., Lewin, G. R., Azpurua, J., Gorbunova, V., Seluanov, A., Regnard, P., Wanert, F., Marchal, J., Pifferi, F., Aujard, F., Liu, Z., Shi, P., Pääbo, S., Schroeder, F., Willmitzer, L., Giavalisco, P., and Khaitovich, P. (2017). Lipidome determinants of maximal lifespan in mammals. *Scientific Reports*, 7(1).
- [Bozek et al., 2014] Bozek, K., Wei, Y., Yan, Z., Liu, X., Xiong, J., Sugimoto, M., Tomita, M., Pääbo, S., Pieszek, R., Sherwood, C. C., Hof, P. R., Ely, J. J., Steinhauser, D., Willmitzer, L., Bangsbo, J., Hansson, O., Call, J., Giavalisco, P., and Khaitovich, P. (2014). Exceptional Evolutionary Divergence of Human Muscle and Brain Metabolomes Parallels Human Cognitive and Physical Uniqueness. *PLoS Biology*, 12(5):e1001871.
- [Bozek et al., 2015] Bozek, K., Wei, Y., Yan, Z., Liu, X., Xiong, J., Sugimoto, M., Tomita, M., Pääbo, S., Sherwood, C., Hof, P., Ely, J., Li, Y., Steinhauser, D., Willmitzer, L., Giavalisco, P., and Khaitovich, P. (2015). Organization and Evolution of Brain Lipidome Revealed by Large-Scale Analysis of Human, Chimpanzee, Macaque, and Mouse Tissues. *Neuron*, 85(4):695–702.

- [Brawand et al., 2011] Brawand, D., Soumillon, M., Necsulea, A., Julien, P., Csárdi, G., Harrigan, P., Weier, M., Liechti, A., Aximu-Petri, A., Kircher, M., Albert, F. W., Zeller, U., Khaitovich, P., Grützner, F., Bergmann, S., Nielsen, R., Pääbo, S., and Kaessmann, H. (2011). The evolution of gene expression levels in mammalian organs. *Nature*, 478(7369):343–348.
- [Breitling et al., 2013] Breitling, R., Cenicerros, A., Jankevics, A., and Takano, E. (2013). Metabolomics for Secondary Metabolite Research. *Metabolites*, 3(4):1076–1083.
- [Brigandi et al., 2015] Brigandi, S., Shao, H., Qian, S., Shen, Y., Wu, B.-L., and Kang, J. (2015). Autistic Children Exhibit Decreased Levels of Essential Fatty Acids in Red Blood Cells. *International Journal of Molecular Sciences*, 16(12):10061–10076.
- [Brimacombe et al., 2007] Brimacombe, M. B., Pickett, R., and Pickett, J. (2007). Autism Post-Mortem Neuroinformatic Resource: The Autism Tissue Program (ATP) Informatics Portal. *Journal of Autism and Developmental Disorders*, 37(3):574–579.
- [Bu et al., 2006] Bu, B., Ashwood, P., Harvey, D., King, I., Water, J. d., and Jin, L.-W. (2006). Fatty acid compositions of red blood cell phospholipids in children with autism. *Prostaglandins, Leukotrienes and Essential Fatty Acids*, 74(4):215–221.
- [Burnstock, 2011] Burnstock, G. (2011). Introductory overview of purinergic signalling. *Frontiers in Bioscience (Elite Edition)*, 3:896–900.
- [Burnstock et al., 2011] Burnstock, G., Krugel, U., Abbracchio, M. P., and Illes, P. (2011). Purinergic signalling: From normal behaviour to pathological brain function. *Progress in Neurobiology*, 95(2):229–274.
- [Butler and King, 2004] Butler, M. and King, A. (2004). Phylogenetic Comparative Analysis: A Modeling Approach for Adaptive Evolution. *The American Naturalist*, 164(6):683–695.
- [Buyske et al., 2006] Buyske, S., Williams, T. A., Mars, A. E., Stenroos, E. S., Ming, S. X., Wang, R., Sreenath, M., Factura, M. F., Reddy, C., Lambert, G. H., and Johnson, W. G. (2006). Analysis of case-parent trios at a locus with a deletion allele: association of GSTM1 with autism. *BMC genetics*, 7:8.
- [Careaga et al., 2013] Careaga, M., Hansen, R. L., Hertz-Piccolto, I., Van de Water, J., and Ashwood, P. (2013). Increased Anti-Phospholipid Antibodies in Autism Spectrum Disorders. *Mediators of Inflammation*, 2013:1–7.
- [Chakrabarti et al., 2015] Chakrabarti, B., Persico, A., Battista, N., and Maccarrone, M. (2015). Endocannabinoid Signaling in Autism. *Neurotherapeutics*, 12(4):837–847.
- [Chan et al., 2012] Chan, R. B., Oliveira, T. G., Cortes, E. P., Honig, L. S., Duff, K. E., Small, S. A., Wenk, M. R., Shui, G., and Di Paolo, G. (2012). Comparative Lipidomic Analysis of Mouse and Human Brain with Alzheimer Disease. *Journal of Biological Chemistry*, 287(4):2678–2688.

- [Chang et al., 2011] Chang, S.-H., Chiang, S.-Y., Chiu, C.-C., Tsai, C.-C., Tsai, H.-H., Huang, C.-Y., Hsu, T.-C., and Tzang, B.-S. (2011). Expression of anti-cardiolipin antibodies and inflammatory associated factors in patients with schizophrenia. *Psychiatry Research*, 187(3):341–346.
- [Cheffer et al., 2018] Cheffer, A., Castillo, A. R. G., Corrêa-Velloso, J., Gonçalves, M. C. B., Naaldijk, Y., Nascimento, I. C., Burnstock, G., and Ulrich, H. (2018). Purinergic system in psychiatric diseases. *Molecular Psychiatry*, 23(1):94–106.
- [Cheng et al., 2011] Cheng, D., Jenner, A. M., Shui, G., Cheong, W. F., Mitchell, T. W., Nealon, J. R., Kim, W. S., McCann, H., Wenk, M. R., Halliday, G. M., and Garner, B. (2011). Lipid Pathway Alterations in Parkinson’s Disease Primary Visual Cortex. *PLoS ONE*, 6(2):e17299.
- [Chiapparino et al., 2016] Chiapparino, A., Maeda, K., Turei, D., Saez-Rodriguez, J., and Gavin, A.-C. (2016). The orchestra of lipid-transfer proteins at the crossroads between metabolism and signaling. *Progress in Lipid Research*, 61:30–39.
- [Chung et al., 2007] Chung, K.-H., Tsai, S.-Y., and Lee, H.-C. (2007). Mood symptoms and serum lipids in acute phase of bipolar disorder in Taiwan. *Psychiatry and Clinical Neurosciences*, 61(4):428–433.
- [Cohen, 2004] Cohen, J. C. (2004). Multiple Rare Alleles Contribute to Low Plasma Levels of HDL Cholesterol. *Science*, 305(5685):869–872.
- [Colsch et al., 2008] Colsch, B., Afonso, C., Turpin, J.-C., Portoukalian, J., Tabet, J.-C., and Baumann, N. (2008). Sulfogalactosylceramides in motor and psycho-cognitive adult metachromatic leukodystrophy: relations between clinical, biochemical analysis and molecular aspects. *Biochimica et Biophysica Acta (BBA) - General Subjects*, 1780(3):434–440.
- [Cotsapas et al., 2011] Cotsapas, C., Voight, B. F., Rossin, E., Lage, K., Neale, B. M., Wallace, C., Abecasis, G. R., Barrett, J. C., Behrens, T., Cho, J., De Jager, P. L., Elder, J. T., Graham, R. R., Gregersen, P., Klareskog, L., Siminovitch, K. A., van Heel, D. A., Wijmenga, C., Worthington, J., Todd, J. A., Hafler, D. A., Rich, S. S., Daly, M. J., and on behalf of the FOCiS Network of Consortia (2011). Pervasive Sharing of Genetic Effects in Autoimmune Disease. *PLoS Genetics*, 7(8):e1002254.
- [Cozzolino et al., 2014] Cozzolino, R., De Magistris, L., Saggese, P., Stocchero, M., Martignetti, A., Di Stasio, M., Malorni, A., Marotta, R., Boscaino, F., and Malorni, L. (2014). Use of solid-phase microextraction coupled to gas chromatography–mass spectrometry for determination of urinary volatile organic compounds in autistic children compared with healthy controls. *Analytical and Bioanalytical Chemistry*, 406(19):4649–4662.
- [Crowder et al., 2017] Crowder, M. K., Seacrist, C. D., and Blind, R. D. (2017). Phospholipid regulation of the nuclear receptor superfamily. *Advances in Biological Regulation*, 63:6–14.
- [De Craene et al., 2017] De Craene, J.-O., Bertazzi, D., Bar, S., and Friant, S. (2017). Phosphoinositides, Major Actors in Membrane Trafficking and Lipid Signaling Pathways. *International Journal of Molecular Sciences*, 18(3):634.

- [de la Torre-Ubieta et al., 2016] de la Torre-Ubieta, L., Won, H., Stein, J. L., and Geschwind, D. H. (2016). Advancing the understanding of autism disease mechanisms through genetics. *Nature Medicine*, 22(4):345–361.
- [De Petrocellis et al., 2000] De Petrocellis, L., Melck, D., Bisogno, T., and Di Marzo, V. (2000). Endocannabinoids and fatty acid amides in cancer, inflammation and related disorders. *Chemistry and Physics of Lipids*, 108(1-2):191–209.
- [De Rubeis and Buxbaum, 2015] De Rubeis, S. and Buxbaum, J. D. (2015). Genetics and genomics of autism spectrum disorder: embracing complexity. *Human Molecular Genetics*, 24(R1):R24–R31.
- [Demain and Fang, 2000] Demain, A. L. and Fang, A. (2000). The natural functions of secondary metabolites. *Advances in Biochemical Engineering/Biotechnology*, 69:1–39.
- [Diémé et al., 2015] Diémé, B., Mavel, S., Blasco, H., Tripi, G., Bonnet-Brilhault, F., Malvy, J., Bocca, C., Andres, C. R., Nadal-Desbarats, L., and Emond, P. (2015). Metabolomics Study of Urine in Autism Spectrum Disorders Using a Multiplatform Analytical Methodology. *Journal of Proteome Research*, 14(12):5273–5282.
- [Domański et al., 2006] Domański, L., Safranow, K., Dolegowska, B., Rózański, J., Myślak, M., Ciechanowski, K., Jakubowska, K., Dziedziejko, V., Romanowski, M., Sulikowski, T., Sieńko, J., Kamiński, M., Ostrowski, M., Domański, M., Pawlik, A., Rać, M. E., and Chlubek, D. (2006). Hypoxanthine as a graft ischemia marker stimulates catalase activity in the renal vein during reperfusion in humans. *Transplantation Proceedings*, 38(1):35–38.
- [Douzery et al., 2014] Douzery, E. J. P., Scornavacca, C., Romiguier, J., Belkhir, K., Galtier, N., Delsuc, F., and Ranwez, V. (2014). OrthoMaM v8: A Database of Orthologous Exons and Coding Sequences for Comparative Genomics in Mammals. *Molecular Biology and Evolution*, 31(7):1923–1928.
- [Dowhan and Bogdanov, 2002] Dowhan, W. and Bogdanov, M. (2002). Chapter 1 Functional roles of lipids in membranes. In *New Comprehensive Biochemistry*, volume 36, pages 1–35. Elsevier.
- [Eapen, 2011] Eapen, V., editor (2011). *Autism - A Neurodevelopmental Journey from Genes to Behaviour*. InTech.
- [Eggan et al., 2008] Eggan, S. M., Hashimoto, T., and Lewis, D. A. (2008). Reduced Cortical Cannabinoid 1 Receptor Messenger RNA and Protein Expression in Schizophrenia. *Archives of General Psychiatry*, 65(7):772.
- [El-Ansary et al., 2017] El-Ansary, A., Bjørklund, G., Chirumbolo, S., and Alnakhli, O. M. (2017). Predictive value of selected biomarkers related to metabolism and oxidative stress in children with autism spectrum disorder. *Metabolic Brain Disease*, 32(4):1209–1221.
- [El-Ansary et al., 2011] El-Ansary, A. K., Ben Bacha, A. G., and Al-Ayahdi, L. Y. (2011). Plasma fatty acids as diagnostic markers in autistic patients from Saudi Arabia. *Lipids in Health and Disease*, 10(1):62.

- [Elphick, 2012] Elphick, M. R. (2012). The evolution and comparative neurobiology of endocannabinoid signalling. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367(1607):3201–3215.
- [Emond et al., 2013] Emond, P., Mavel, S., Aïdoud, N., Nadal-Desbarats, L., Montigny, F., Bonnet-Brilhault, F., Barthélémy, C., Merten, M., Sarda, P., Laumonnier, F., Vourc’h, P., Blasco, H., and Andres, C. R. (2013). GC-MS-based urine metabolic profiling of autism spectrum disorders. *Analytical and Bioanalytical Chemistry*, 405(15):5291–5300.
- [Ernst et al., 2016] Ernst, R., Ejsing, C. S., and Antonny, B. (2016). Homeoviscous Adaptation and the Regulation of Membrane Lipids. *Journal of Molecular Biology*, 428(24):4776–4791.
- [Eyre-Walker, 2006] Eyre-Walker, A. (2006). The genomic rate of adaptive evolution. *Trends in Ecology & Evolution*, 21(10):569–575.
- [Fahy et al., 2009] Fahy, E., Subramaniam, S., Murphy, R. C., Nishijima, M., Raetz, C. R. H., Shimizu, T., Spener, F., van Meer, G., Wakelam, M. J. O., and Dennis, E. A. (2009). Update of the LIPID MAPS comprehensive classification system for lipids. *Journal of Lipid Research*, 50(Supplement):S9–S14.
- [Filipek et al., 2004] Filipek, P. A., Juranek, J., Nguyen, M. T., Cummings, C., and Gargus, J. J. (2004). Relative Carnitine Deficiency in Autism. *Journal of Autism and Developmental Disorders*, 34(6):615–623.
- [Frazer et al., 2009] Frazer, K. A., Murray, S. S., Schork, N. J., and Topol, E. J. (2009). Human genetic variation and its contribution to complex traits. *Nature Reviews Genetics*, 10(4):241–251.
- [Fu et al., 2011] Fu, X., Giavalisco, P., Liu, X., Catchpole, G., Fu, N., Ning, Z.-B., Guo, S., Yan, Z., Somel, M., Paabo, S., Zeng, R., Willmitzer, L., and Khaitovich, P. (2011). Rapid metabolic evolution in human prefrontal cortex. *Proceedings of the National Academy of Sciences*, 108(15):6181–6186.
- [Fushan et al., 2015] Fushan, A. A., Turanov, A. A., Lee, S.-G., Kim, E. B., Lobanov, A. V., Yim, S. H., Buffenstein, R., Lee, S.-R., Chang, K.-T., Rhee, H., Kim, J.-S., Yang, K.-S., and Gladyshev, V. N. (2015). Gene expression defines natural changes in mammalian lifespan. *Aging Cell*, 14(3):352–365.
- [Garber et al., 1976] Garber, A. J., Karl, I. E., and Kipnis, D. M. (1976). Alanine and glutamine synthesis and release from skeletal muscle. I. Glycolysis and amino acid release. *The Journal of Biological Chemistry*, 251(3):826–835.
- [Gennis, 1989] Gennis, R. B. (1989). *Biomembranes: Molecular Structure and Function*. Springer New York, New York, NY. OCLC: 853257926.
- [Gevi et al., 2016] Gevi, F., Zolla, L., Gabriele, S., and Persico, A. M. (2016). Urinary metabolomics of young Italian autistic children supports abnormal tryptophan and purine metabolism. *Molecular Autism*, 7(1).

- [Giavalisco et al., 2011] Giavalisco, P., Li, Y., Matthes, A., Eckhardt, A., Hubberten, H.-M., Hesse, H., Segu, S., Hummel, J., Köhl, K., and Willmitzer, L. (2011). Elemental formula annotation of polar and lipophilic metabolites using ^{13}C , ^{15}N and ^{34}S isotope labelling, in combination with high-resolution mass spectrometry: *Isotope labelling for unbiased plant metabolomics*. *The Plant Journal*, 68(2):364–376.
- [Gordon Bell et al., 2009] Gordon Bell, J., Miller, D., MacDonald, D. J., MacKinlay, E. E., Dick, J. R., Cheseldine, S., Boyle, R. M., Graham, C., and O’Hare, A. E. (2009). The fatty acid compositions of erythrocyte and plasma polar lipids in children with autism, developmental delay or typically developing controls and the effect of fish oil intake. *British Journal of Nutrition*, page 1.
- [Graham et al., 2016] Graham, S. F., Chevallier, O. P., Kumar, P., Türkoğlu, O., and Bahado-Singh, R. O. (2016). High resolution metabolomic analysis of ASD human brain uncovers novel biomarkers of disease. *Metabolomics*, 12(4).
- [Grösch et al., 2012] Grösch, S., Schiffmann, S., and Geisslinger, G. (2012). Chain length-specific properties of ceramides. *Progress in Lipid Research*, 51(1):50–62.
- [Gubert et al., 2016] Gubert, C., Jacintho Moritz, C. E., Vasconcelos-Moreno, M. P., Quadros dos Santos, B. T. M., Sartori, J., Fijtman, A., Kauer-Sant’Anna, M., Kapczinski, F., Battastini, A. M. O., and Magalhães, P. V. d. S. (2016). Peripheral adenosine levels in euthymic patients with bipolar disorder. *Psychiatry Research*, 246:421–426.
- [Hahn, 2008] Hahn, M. W. (2008). Toward a selection theory of molecular evolution. *Evolution; International Journal of Organic Evolution*, 62(2):255–265.
- [Han, 2016] Han, X. (2016). *Lipidomics: comprehensive mass spectrometry of lipids*. Wiley, Hoboken, New Jersey.
- [Han et al., 2002] Han, X., M Holtzman, D., McKeel, D. W., Kelley, J., and Morris, J. C. (2002). Substantial sulfatide deficiency and ceramide elevation in very early Alzheimer’s disease: potential role in disease pathogenesis. *Journal of Neurochemistry*, 82(4):809–818.
- [Harayama and Riezman, 2018] Harayama, T. and Riezman, H. (2018). Understanding the diversity of membrane lipid composition. *Nature Reviews Molecular Cell Biology*, 19(5):281–296.
- [Haughey et al., 2010] Haughey, N. J., Bandaru, V. V., Bae, M., and Mattson, M. P. (2010). Roles for dysfunctional sphingolipid metabolism in Alzheimer’s disease neuropathogenesis. *Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids*, 1801(8):878–886.
- [Heimburg, 2007] Heimburg, T. (2007). *Thermal Biophysics of Membranes*. Tutorials in Biophysics. Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, Germany.
- [Helguera et al., 2013] Helguera, P., Seiglie, J., Rodriguez, J., Hanna, M., Helguera, G., and Busciglio, J. (2013). Adaptive Downregulation of Mitochondrial Function in Down Syndrome. *Cell Metabolism*, 17(1):132–140.

- [Hicks et al., 2006] Hicks, A. M., DeLong, C. J., Thomas, M. J., Samuel, M., and Cui, Z. (2006). Unique molecular signatures of glycerophospholipid species in different rat tissues analyzed by tandem mass spectrometry. *Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids*, 1761(9):1022–1029.
- [Horwege et al., 2014] Horwege, S., Lindner, S., Boden, M., Hatje, K., Kollmar, M., Leimeister, C.-A., and Morgenstern, B. (2014). Spaced words and kmacs: fast alignment-free sequence comparison based on inexact word matches. *Nucleic Acids Research*, 42(W1):W7–W11.
- [Ingram and Mahler, 2013] Ingram, T. and Mahler, D. (2013). SURFACE: detecting convergent evolution from comparative data by fitting Ornstein-Uhlenbeck models with stepwise Akaike Information Criterion. *Methods in Ecology and Evolution*, 4(5):416–425.
- [James, 2017] James, K.-D. (2017). Animal Metabolites. In *Pharmacognosy*, pages 401–411. Elsevier.
- [James et al., 2004] James, S. J., Cutler, P., Melnyk, S., Jernigan, S., Janak, L., Gaylor, D. W., and Neubrandner, J. A. (2004). Metabolic biomarkers of increased oxidative stress and impaired methylation capacity in children with autism. *The American Journal of Clinical Nutrition*, 80(6):1611–1617.
- [Jordan et al., 2005] Jordan, I. K., Mariño-Ramírez, L., and Koonin, E. V. (2005). Evolutionary significance of gene expression divergence. *Gene*, 345(1):119–126.
- [Kaddurah-Daouk et al., 2007] Kaddurah-Daouk, R., McEvoy, J., Baillie, R. A., Lee, D., Yao, J. K., Doraiswamy, P. M., and Krishnan, K. R. R. (2007). Metabolomic mapping of atypical antipsychotic effects in schizophrenia. *Molecular Psychiatry*, 12(10):934–945.
- [Kanehisa et al., 2017] Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y., and Morishima, K. (2017). KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Research*, 45(D1):D353–D361.
- [Karhson et al., 2016] Karhson, D. S., Hardan, A. Y., and Parker, K. J. (2016). Endocannabinoid signaling in social functioning: an RDoC perspective. *Translational Psychiatry*, 6(9):e905–e905.
- [Kałużna-Czaplińska et al., 2014] Kałużna-Czaplińska, J., Żurawicz, E., Struck, W., and Markuszewski, M. (2014). Identification of organic acids as potential biomarkers in the urine of autistic children using gas chromatography/mass spectrometry. *Journal of Chromatography B*, 966:70–76.
- [Khaitovich et al., 2008] Khaitovich, P., Lockstone, H. E., Wayland, M. T., Tsang, T. M., Jayatilaka, S. D., Guo, A. J., Zhou, J., Somel, M., Harris, L. W., Holmes, E., Pääbo, S., and Bahn, S. (2008). Metabolic changes in schizophrenia and human brain evolution. *Genome Biology*, 9(8):R124.
- [Khrameeva et al., 2018] Khrameeva, E., Kurochkin, I., Bozek, K., Giavalisco, P., and Khaitovich, P. (2018). Lipidome Evolution in Mammalian Tissues. *Molecular Biology and Evolution*, 35(8):1947–1957.

- [Kim et al., 2010] Kim, E.-K., Neggers, Y. H., Shin, C.-S., Kim, E., and Kim, E. M. (2010). Alterations in lipid profile of autistic boys: a case control study. *Nutrition Research*, 30(4):255–260.
- [Kumar et al., 2017] Kumar, S., Stecher, G., Suleski, M., and Hedges, S. B. (2017). Time-Tree: A Resource for Timelines, Timetrees, and Divergence Times. *Molecular Biology and Evolution*, 34(7):1812–1819.
- [Kuraku et al., 2016] Kuraku, S., Feiner, N., Keeley, S. D., and Hara, Y. (2016). Incorporating tree-thinking and evolutionary time scale into developmental biology. *Development, Growth & Differentiation*, 58(1):131–142.
- [Kuratko and Salem, 2009] Kuratko, C. N. and Salem, N. (2009). Biomarkers of DHA status. *Prostaglandins, Leukotrienes and Essential Fatty Acids*, 81(2-3):111–118.
- [Lamari et al., 2013] Lamari, F., Mochel, F., Sedel, F., and Saudubray, J. M. (2013). Disorders of phospholipids, sphingolipids and fatty acids biosynthesis: toward a new category of inherited metabolic diseases. *Journal of Inherited Metabolic Disease*, 36(3):411–425.
- [Lanfear et al., 2010] Lanfear, R., Welch, J. J., and Bromham, L. (2010). Watching the clock: Studying variation in rates of molecular evolution between species. *Trends in Ecology & Evolution*, 25(9):495–503.
- [Laszlo et al., 1994] Laszlo, A., Horváth, E., Eck, E., and Fekete, M. (1994). Serum serotonin, lactate and pyruvate levels in infantile autistic children. *Clinica Chimica Acta*, 229(1-2):205–207.
- [Lauwers et al., 2016] Lauwers, E., Goodchild, R., and Verstreken, P. (2016). Membrane Lipids in Presynaptic Function and Disease. *Neuron*, 90(1):11–25.
- [Leslie et al., 2014] Leslie, R., O’Donnell, C. J., and Johnson, A. D. (2014). GRASP: analysis of genotype-phenotype results from 1390 genome-wide association studies and corresponding open access database. *Bioinformatics*, 30(12):i185–i194.
- [Li et al., 2017] Li, Q., Bozek, K., Xu, C., Guo, Y., Sun, J., Pääbo, S., Sherwood, C. C., Hof, P. R., Ely, J. J., Li, Y., Willmitzer, L., Giavalisco, P., and Khaitovich, P. (2017). Changes in Lipidome Composition during Brain Development in Humans, Chimpanzees, and Macaque Monkeys. *Molecular Biology and Evolution*, 34(5):1155–1166.
- [Li et al., 2016] Li, Y. I., van de Geijn, B., Raj, A., Knowles, D. A., Petti, A. A., Golan, D., Gilad, Y., and Pritchard, J. K. (2016). RNA splicing is a primary link between genetic variation and disease. *Science*, 352(6285):600–604.
- [Liu et al., 2016] Liu, X., Han, D., Somel, M., Jiang, X., Hu, H., Guijarro, P., Zhang, N., Mitchell, A., Halene, T., Ely, J. J., Sherwood, C. C., Hof, P. R., Qiu, Z., Pääbo, S., Akbarian, S., and Khaitovich, P. (2016). Disruption of an Evolutionarily Novel Synaptic Expression Pattern in Autism. *PLOS Biology*, 14(9):e1002558.
- [Loffler et al., 2005] Loffler, M., Fairbanks, L., Zameitat, E., Marinaki, A., and Simmonds, H. (2005). Pyrimidine pathways in health and disease. *Trends in Molecular Medicine*, 11(9):430–437.

- [Lombard, 2006] Lombard, J. 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*, 290(3):H921–922.
- [Lussu et al., 2017] Lussu, M., Noto, A., Masili, A., Rinaldi, A. C., Dessì, A., De Angelis, M., De Giacomo, A., Fanos, V., Atzori, L., and Francavilla, R. (2017). The urinary ^1H -NMR metabolomics profile of an italian autistic children population and their unaffected siblings: Metabolomics profile of autistic children. *Autism Research*, 10(6):1058–1066.
- [Ma et al., 2015] Ma, S., Yim, S., Lee, S.-G., Kim, E., Lee, S.-R., Chang, K.-T., Buffenstein, R., Lewis, K., Park, T., Miller, R., Clish, C., and Gladyshev, V. (2015). Organization of the Mammalian Metabolome according to Organ Function, Lineage Specialization, and Longevity. *Cell Metabolism*, 22(2):332–343.
- [Mavel et al., 2013] Mavel, S., Nadal-Desbarats, L., Blasco, H., Bonnet-Brilhault, F., Barthelemy, C., Montigny, F., Sarda, P., Laumonnier, F., Vourch, P., Andres, C. R., and Emond, P. (2013). ^1H - ^{13}C NMR-based urine metabolic profiling in autism spectrum disorders. *Talanta*, 114:95–102.
- [McEvoy et al., 2013] McEvoy, J., Baillie, R. A., Zhu, H., Buckley, P., Keshavan, M. S., Nasrallah, H. A., Dougherty, G. G., Yao, J. K., and Kaddurah-Daouk, R. (2013). Lipidomics Reveals Early Metabolic Changes in Subjects with Schizophrenia: Effects of Atypical Antipsychotics. *PLoS ONE*, 8(7):e68717.
- [McPartland et al., 2006] McPartland, J. M., Matias, I., Di Marzo, V., and Glass, M. (2006). Evolutionary origins of the endocannabinoid system. *Gene*, 370:64–74.
- [Meinshausen and Buehlmann, 2008] Meinshausen, N. and Buehlmann, P. (2008). Stability Selection. *arXiv:0809.2932 [stat]*. arXiv: 0809.2932.
- [Mergenthaler et al., 2013] Mergenthaler, P., Lindauer, U., Dienel, G. A., and Meisel, A. (2013). Sugar for the brain: the role of glucose in physiological and pathological brain function. *Trends in Neurosciences*, 36(10):587–597.
- [Micheli et al., 2011] Micheli, V., Camici, M., G. Tozzi, M., L. Ipata, P., Sestini, S., Bertelli, M., and Pompucci, G. (2011). Neurological Disorders of Purine and Pyrimidine Metabolism. *Current Topics in Medicinal Chemistry*, 11(8):923–947.
- [Ming et al., 2010] Ming, X., Johnson, W. G., Stenroos, E. S., Mars, A., Lambert, G. H., and Buyske, S. (2010). Genetic variant of glutathione peroxidase 1 in autism. *Brain and Development*, 32(2):105–109.
- [Ming et al., 2012] Ming, X., Stein, T. P., Barnes, V., Rhodes, N., and Guo, L. (2012). Metabolic Perturbance in Autism Spectrum Disorders: A Metabolomics Study. *Journal of Proteome Research*, 11(12):5856–5862.
- [Morgan et al., 2006] Morgan, N. V., Westaway, S. K., Morton, J. E. V., Gregory, A., Gissen, P., Sonek, S., Cangul, H., Coryell, J., Canham, N., Nardocci, N., Zorzi, G., Pasha, S., Rodriguez, D., Desguerre, I., Mubaidin, A., Bertini, E., Trembath, R. C., Simonati,

- A., Schanen, C., Johnson, C. A., Levinson, B., Woods, C. G., Wilmot, B., Kramer, P., Gitschier, J., Maher, E. R., and Hayflick, S. J. (2006). PLA2g6, encoding a phospholipase A2, is mutated in neurodegenerative disorders with high brain iron. *Nature Genetics*, 38(7):752–754.
- [Muguruza et al., 2013] Muguruza, C., Lehtonen, M., Aaltonen, N., Morentin, B., Meana, J. J., and Callado, L. F. (2013). Quantification of endocannabinoids in postmortem brain of schizophrenic subjects. *Schizophrenia Research*, 148(1-3):145–150.
- [Nadal-Desbarats et al., 2014] Nadal-Desbarats, L., Aidoud, N., Emond, P., Blasco, H., Filipiak, I., Sarda, P., Bonnet-Brilhault, F., Mavel, S., and Andres, C. R. (2014). Combined ^1H -NMR and ^1H - ^{13}C HSQC-NMR to improve urinary screening in autism spectrum disorders. *The Analyst*, 139(13):3460–3468.
- [Nei, 2007] Nei, M. (2007). The new mutation theory of phenotypic evolution. *Proceedings of the National Academy of Sciences of the United States of America*, 104(30):12235–12242.
- [Noto et al., 2014] Noto, A., Fanos, V., Barberini, L., Grapov, D., Fattuoni, C., Zaffanello, M., Casanova, A., Fenu, G., De Giacomo, A., De Angelis, M., Moretti, C., Papoff, P., Ditunno, R., and Francavilla, R. (2014). The urinary metabolomics profile of an Italian autistic children population and their unaffected siblings. *The Journal of Maternal-Fetal & Neonatal Medicine*, 27(sup2):46–52.
- [Orešič et al., 2012] Orešič, M., Seppänen-Laakso, T., Sun, D., Tang, J., Therman, S., Viehman, R., Mustonen, U., van Erp, T. G., Hyötyläinen, T., Thompson, P., Toga, A. W., Huttunen, M. O., Suvisaari, J., Kaprio, J., Lönngqvist, J., and Cannon, T. D. (2012). Phospholipids and insulin resistance in psychosis: a lipidomics study of twin pairs discordant for schizophrenia. *Genome Medicine*, 4(1):1.
- [Orr, 2009] Orr, H. A. (2009). Fitness and its role in evolutionary genetics. *Nature Reviews. Genetics*, 10(8):531–539.
- [Otto, 2000] Otto, S. P. (2000). Detecting the form of selection from DNA sequence data. *Trends in genetics: TIG*, 16(12):526–529.
- [Pagel, 1999] Pagel, M. (1999). Inferring the historical patterns of biological evolution. *Nature*, 401(6756):877–884.
- [Parletta et al., 2016] Parletta, N., Niyonsenga, T., and Duff, J. (2016). Omega-3 and Omega-6 Polyunsaturated Fatty Acid Levels and Correlations with Symptoms in Children with Attention Deficit Hyperactivity Disorder, Autistic Spectrum Disorder and Typically Developing Controls. *PLOS ONE*, 11(5):e0156432.
- [Pastural et al., 2009] Pastural, E., Ritchie, S., Lu, Y., Jin, W., Kavianpour, A., Khine Su-Myat, K., Heath, D., Wood, P. L., Fisk, M., and Goodenowe, D. B. (2009). Novel plasma phospholipid biomarkers of autism: Mitochondrial dysfunction as a putative causative mechanism. *Prostaglandins, Leukotrienes and Essential Fatty Acids*, 81(4):253–264.

- [Pazos et al., 2008] Pazos, M., Sagredo, O., and Fernandez-Ruiz, J. (2008). The Endocannabinoid System in Huntingtons Disease. *Current Pharmaceutical Design*, 14(23):2317–2325.
- [Peter, 2004] Peter, B. J. (2004). BAR Domains as Sensors of Membrane Curvature: The Amphiphysin BAR Structure. *Science*, 303(5657):495–499.
- [Pickrell et al., 2016] Pickrell, J. K., Berisa, T., Liu, J. Z., Ségurel, L., Tung, J. Y., and Hinds, D. A. (2016). Detection and interpretation of shared genetic influences on 42 human traits. *Nature Genetics*, 48(7):709–717.
- [Piomelli, 2003] Piomelli, D. (2003). The molecular logic of endocannabinoid signalling. *Nature Reviews Neuroscience*, 4(11):873–884.
- [Piomelli et al., 2007] Piomelli, D., Astarita, G., and Rapaka, R. (2007). A neuroscientist’s guide to lipidomics. *Nature Reviews Neuroscience*, 8(10):743–754.
- [Psychogios et al., 2011] Psychogios, N., Hau, D. D., Peng, J., Guo, A. C., Mandal, R., Bouatra, S., Sinelnikov, I., Krishnamurthy, R., Eisner, R., Gautam, B., Young, N., Xia, J., Knox, C., Dong, E., Huang, P., Hollander, Z., Pedersen, T. L., Smith, S. R., Bamforth, F., Greiner, R., McManus, B., Newman, J. W., Goodfriend, T., and Wishart, D. S. (2011). The Human Serum Metabolome. *PLoS ONE*, 6(2):e16957.
- [Pérez-Rodríguez et al., 2015] Pérez-Rodríguez, L., Romero-Haro, A. A., Sternalski, A., Muriel, J., Mougeot, F., Gil, D., and Alonso-Alvarez, C. (2015). Measuring Oxidative Stress: The Confounding Effect of Lipid Concentration in Measures of Lipid Peroxidation. *Physiological and Biochemical Zoology*, 88(3):345–351.
- [ReproGen Consortium et al., 2015] ReproGen Consortium, Psychiatric Genomics Consortium, Genetic Consortium for Anorexia Nervosa of the Wellcome Trust Case Control Consortium 3, Bulik-Sullivan, B., Finucane, H. K., Anttila, V., Gusev, A., Day, F. R., Loh, P.-R., Duncan, L., Perry, J. R. B., Patterson, N., Robinson, E. B., Daly, M. J., Price, A. L., and Neale, B. M. (2015). An atlas of genetic correlations across human diseases and traits. *Nature Genetics*, 47(11):1236–1241.
- [Rose et al., 2012] Rose, S., Melnyk, S., Pavliv, O., Bai, S., Nick, T. G., Frye, R. E., and James, S. J. (2012). Evidence of oxidative damage and inflammation associated with low glutathione redox status in the autism brain. *Translational Psychiatry*, 2(7):e134–e134.
- [Rossi et al., 2012] Rossi, L., Salvestrini, V., Ferrari, D., Di Virgilio, F., and Lemoli, R. M. (2012). The sixth sense: hematopoietic stem cells detect danger through purinergic signaling. *Blood*, 120(12):2365–2375.
- [Rossignol and Frye, 2012] Rossignol, D. A. and Frye, R. E. (2012). Mitochondrial dysfunction in autism spectrum disorders: a systematic review and meta-analysis. *Molecular Psychiatry*, 17(3):290–314.
- [Sanders et al., 2012] Sanders, S. J., Murtha, M. T., Gupta, A. R., Murdoch, J. D., Raubeson, M. J., Willsey, A. J., Ercan-Sencicek, A. G., DiLullo, N. M., Parikshak, N. N., Stein,

- J. L., Walker, M. F., Ober, G. T., Teran, N. A., Song, Y., El-Fishawy, P., Murtha, R. C., Choi, M., Overton, J. D., Bjornson, R. D., Carriero, N. J., Meyer, K. A., Bilguvar, K., Mane, S. M., Šestan, N., Lifton, R. P., Günel, M., Roeder, K., Geschwind, D. H., Devlin, B., and State, M. W. (2012). De novo mutations revealed by whole-exome sequencing are strongly associated with autism. *Nature*, 485(7397):237–241.
- [Sandin et al., 2016] Sandin, S., Schendel, D., Magnusson, P., Hultman, C., Surén, P., Susser, E., Grønberg, T., Gissler, M., Gunnes, N., Gross, R., Henning, M., Bresnahan, M., Sourander, A., Hornig, M., Carter, K., Francis, R., Parner, E., Leonard, H., Rosanoff, M., Stoltenberg, C., and Reichenberg, A. (2016). Autism risk associated with parental age and with increasing difference in age between the parents. *Molecular Psychiatry*, 21(5):693–700.
- [Schmitt et al., 2004] Schmitt, A., Wilczek, K., Blennow, K., Maras, A., Jatzko, A., Petroianu, G., Braus, D. F., and Gattaz, W. F. (2004). Altered thalamic membrane phospholipids in schizophrenia: a postmortem study. *Biological Psychiatry*, 56(1):41–45.
- [Schwarz et al., 2008] Schwarz, E., Prabakaran, S., Whitfield, P., Major, H., Leweke, F. M., Koethe, D., McKenna, P., and Bahn, S. (2008). High Throughput Lipidomic Profiling of Schizophrenia and Bipolar Disorder Brain Tissue Reveals Alterations of Free Fatty Acids, Phosphatidylcholines, and Ceramides. *Journal of Proteome Research*, 7(10):4266–4277.
- [Sethi et al., 2017] Sethi, S., Hayashi, M. A. F., Barbosa, B. S., Pontes, J. G. M., Tasic, L., and Brietzke, E. (2017). Lipidomics, Biomarkers, and Schizophrenia: A Current Perspective. In Sussulini, A., editor, *Metabolomics: From Fundamentals to Clinical Applications*, volume 965, pages 265–290. Springer International Publishing, Cham.
- [Shao et al., 2008] Shao, H., Burrage, L. C., Sinasac, D. S., Hill, A. E., Ernest, S. R., O’Brien, W., Courtland, H.-W., Jepsen, K. J., Kirby, A., Kulbokas, E. J., Daly, M. J., Broman, K. W., Lander, E. S., and Nadeau, J. H. (2008). Genetic architecture of complex traits: Large phenotypic effects and pervasive epistasis. *Proceedings of the National Academy of Sciences*, 105(50):19910–19914.
- [Sherman and Brophy, 2005] Sherman, D. L. and Brophy, P. J. (2005). Mechanisms of axon ensheathment and myelin growth. *Nature Reviews Neuroscience*, 6(9):683–690.
- [Shimmura et al., 2011] Shimmura, C., Suda, S., Tsuchiya, K. J., Hashimoto, K., Ohno, K., Matsuzaki, H., Iwata, K., Matsumoto, K., Wakuda, T., Kamenno, Y., Suzuki, K., Tsujii, M., Nakamura, K., Takei, N., and Mori, N. (2011). Alteration of Plasma Glutamate and Glutamine Levels in Children with High-Functioning Autism. *PLoS ONE*, 6(10):e25340.
- [Shinohe et al., 2006] Shinohe, A., Hashimoto, K., Nakamura, K., Tsujii, M., Iwata, Y., Tsuchiya, K. J., Sekine, Y., Suda, S., Suzuki, K., Sugihara, G.-i., Matsuzaki, H., Minabe, Y., Sugiyama, T., Kawai, M., Iyo, M., Takei, N., and Mori, N. (2006). Increased serum levels of glutamate in adult patients with autism. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, 30(8):1472–1477.
- [Simons and Toomre, 2000] Simons, K. and Toomre, D. (2000). Lipid rafts and signal transduction. *Nature Reviews Molecular Cell Biology*, 1(1):31–39.

- [Snowden et al., 2017] Snowden, S. G., Ebshiana, A. A., Hye, A., An, Y., Pletnikova, O., O'Brien, R., Troncoso, J., Legido-Quigley, C., and Thambisetty, M. (2017). Association between fatty acid metabolism in the brain and Alzheimer disease neuropathology and cognitive performance: A nontargeted metabolomic study. *PLOS Medicine*, 14(3):e1002266.
- [Solberg et al., 2016] Solberg, D. K., Bentsen, H., Refsum, H., and Andreassen, O. A. (2016). Lipid profiles in schizophrenia associated with clinical traits: a five year follow-up study. *BMC Psychiatry*, 16(1).
- [Somel et al., 2009] Somel, M., Franz, H., Yan, Z., Lorenc, A., Guo, S., Giger, T., Kelso, J., Nickel, B., Dannemann, M., Bahn, S., Webster, M. J., Weickert, C. S., Lachmann, M., Paabo, S., and Khaitovich, P. (2009). Transcriptional neoteny in the human brain. *Proceedings of the National Academy of Sciences*, 106(14):5743–5748.
- [Steiner and Tuljapurkar, 2012] Steiner, U. K. and Tuljapurkar, S. (2012). Neutral theory for life histories and individual variability in fitness components. *Proceedings of the National Academy of Sciences of the United States of America*, 109(12):4684–4689.
- [Stern and Orgogozo, 2008] Stern, D. L. and Orgogozo, V. (2008). The loci of evolution: how predictable is genetic evolution? *Evolution; International Journal of Organic Evolution*, 62(9):2155–2177.
- [Stumvoll, 1998] Stumvoll, M. (1998). Glucose production by the human kidney—its importance has been underestimated. *Nephrology, Dialysis, Transplantation: Official Publication of the European Dialysis and Transplant Association - European Renal Association*, 13(12):2996–2999.
- [Tessier et al., 2016] Tessier, C., Sweers, K., Frajerman, A., Bergaoui, H., Ferreri, F., Delva, C., Lapidus, N., Lamaziere, A., Roiser, J. P., De Hert, M., and Nuss, P. (2016). Membrane lipidomics in schizophrenia patients: a correlational study with clinical and cognitive manifestations. *Translational Psychiatry*, 6(10):e906–e906.
- [The DDD Study et al., 2014] The DDD Study, Homozygosity Mapping Collaborative for Autism, UK10K Consortium, The Autism Sequencing Consortium, De Rubeis, S., He, X., Goldberg, A. P., Poultney, C. S., Samocha, K., Ercument Cicek, A., Kou, Y., Liu, L., Fromer, M., Walker, S., Singh, T., Klei, L., Kosmicki, J., Fu, S.-C., Aleksic, B., Biscaldi, M., Bolton, P. F., Brownfeld, J. M., Cai, J., Campbell, N. G., Carracedo, A., Chahrour, M. H., Chiochetti, A. G., Coon, H., Crawford, E. L., Crooks, L., Curran, S. R., Dawson, G., Duketis, E., Fernandez, B. A., Gallagher, L., Geller, E., Guter, S. J., Sean Hill, R., Ionita-Laza, I., Jimenez Gonzalez, P., Kilpinen, H., Klauck, S. M., Klevzon, A., Lee, I., Lei, J., Lehtimäki, T., Lin, C.-F., Ma'ayan, A., Marshall, C. R., McInnes, A. L., Neale, B., Owen, M. J., Ozaki, N., Parellada, M., Parr, J. R., Purcell, S., Puura, K., Rajagopalan, D., Rehnström, K., Reichenberg, A., Sabo, A., Sachse, M., Sanders, S. J., Schafer, C., Schulte-Rüther, M., Skuse, D., Stevens, C., Szatmari, P., Tammimies, K., Valladares, O., Voran, A., Wang, L.-S., Weiss, L. A., Jeremy Willsey, A., Yu, T. W., Yuen, R. K. C., Cook, E. H., Freitag, C. M., Gill, M., Hultman, C. M., Lehner, T., Palotie, A., Schellenberg, G. D., Sklar, P., State, M. W., Sutcliffe, J. S., Walsh, C. A., Scherer, S. W., Zwick, M. E., Barrett, J. C., Cutler, D. J., Roeder, K., Devlin, B., Daly, M. J., and Buxbaum,

- J. D. (2014). Synaptic, transcriptional and chromatin genes disrupted in autism. *Nature*, 515(7526):209–215.
- [the ENGAGE Consortium et al., 2009] the ENGAGE Consortium, Aulchenko, Y. S., Ripatti, S., Lindqvist, I., Boomsma, D., Heid, I. M., Pramstaller, P. P., Penninx, B. W. J. H., Janssens, A. C. J. W., Wilson, J. F., Spector, T., Martin, N. G., Pedersen, N. L., Kyvik, K. O., Kaprio, J., Hofman, A., Freimer, N. B., Jarvelin, M.-R., Gyllensten, U., Campbell, H., Rudan, I., Johansson, A., Marroni, F., Hayward, C., Vitart, V., Jonasson, I., Pattaro, C., Wright, A., Hastie, N., Pichler, I., Hicks, A. A., Falchi, M., Willemsen, G., Hottenga, J.-J., de Geus, E. J. C., Montgomery, G. W., Whitfield, J., Magnusson, P., Saharinen, J., Perola, M., Silander, K., Isaacs, A., Sijbrands, E. J. G., Uitterlinden, A. G., Wittteman, J. C. M., Oostra, B. A., Elliott, P., Ruukonen, A., Sabatti, C., Gieger, C., Meitinger, T., Kronenberg, F., Döring, A., Wichmann, H.-E., Smit, J. H., McCarthy, M. I., van Duijn, C. M., and Peltonen, L. (2009). Loci influencing lipid levels and coronary heart disease risk in 16 European population cohorts. *Nature Genetics*, 41(1):47–55.
- [van Meer et al., 2008] van Meer, G., Voelker, D. R., and Feigenson, G. W. (2008). Membrane lipids: where they are and how they behave. *Nature Reviews Molecular Cell Biology*, 9(2):112–124.
- [Vancassel et al., 2001] Vancassel, S., Durand, G., Barthélémy, C., Lejeune, B., Martineau, J., Guilloteau, D., Andrès, C., and Chalon, S. (2001). Plasma fatty acid levels in autistic children. *Prostaglandins, Leukotrienes and Essential Fatty Acids (PLEFA)*, 65(1):1–7.
- [Voineagu et al., 2011] Voineagu, I., Wang, X., Johnston, P., Lowe, J. K., Tian, Y., Horvath, S., Mill, J., Cantor, R. M., Blencowe, B. J., and Geschwind, D. H. (2011). Transcriptomic analysis of autistic brain reveals convergent molecular pathology. *Nature*, 474(7351):380–384.
- [Vonkova et al., 2015] Vonkova, I., Saliba, A.-E., Deghou, S., Anand, K., Ceschia, S., Doerks, T., Galih, A., Kugler, K., Maeda, K., Rybin, V., van Noort, V., Ellenberg, J., Bork, P., and Gavin, A.-C. (2015). Lipid Cooperativity as a General Membrane-Recruitment Principle for PH Domains. *Cell Reports*, 12(9):1519–1530.
- [Wang et al., 2016] Wang, H., Liang, S., Wang, M., Gao, J., Sun, C., Wang, J., Xia, W., Wu, S., Sumner, S. J., Zhang, F., Sun, C., and Wu, L. (2016). Potential serum biomarkers from a metabolomics study of autism. *Journal of Psychiatry & Neuroscience*, 41(1):27–37.
- [Wei et al., 2017] Wei, D., Allsop, S., Tye, K., and Piomelli, D. (2017). Endocannabinoid Signaling in the Control of Social Behavior. *Trends in Neurosciences*, 40(7):385–396.
- [Wenk, 2005] Wenk, M. R. (2005). Erratum: The emerging field of lipidomics. *Nature Reviews Drug Discovery*, 4(7):594–610.
- [West et al., 2014] West, P. R., Amaral, D. G., Bais, P., Smith, A. M., Egnash, L. A., Ross, M. E., Palmer, J. A., Fontaine, B. R., Conard, K. R., Corbett, B. A., Cezar, G. G., Donley, E. L. R., and Burrier, R. E. (2014). Metabolomics as a Tool for Discovery of Biomarkers of Autism Spectrum Disorder in the Blood Plasma of Children. *PLoS ONE*, 9(11):e112445.

- [Wiest et al., 2009] Wiest, M., German, J., Harvey, D., Watkins, S., and Hertz-Picciotto, I. (2009). Plasma fatty acid profiles in autism: A case-control study. *Prostaglandins, Leukotrienes and Essential Fatty Acids*, 80(4):221–227.
- [Willer et al., 2008] Willer, C. J., Sanna, S., Jackson, A. U., Scuteri, A., Bonnycastle, L. L., Clarke, R., Heath, S. C., Timpson, N. J., Najjar, S. S., Stringham, H. M., Strait, J., Duren, W. L., Maschio, A., Busonero, F., Mulas, A., Albai, G., Swift, A. J., Morken, M. A., Narisu, N., Bennett, D., Parish, S., Shen, H., Galan, P., Meneton, P., Hercberg, S., Zelenika, D., Chen, W.-M., Li, Y., Scott, L. J., Scheet, P. A., Sundvall, J., Watanabe, R. M., Nagaraja, R., Ebrahim, S., Lawlor, D. A., Ben-Shlomo, Y., Davey-Smith, G., Shuldiner, A. R., Collins, R., Bergman, R. N., Uda, M., Tuomilehto, J., Cao, A., Collins, F. S., Lakatta, E., Lathrop, G. M., Boehnke, M., Schlessinger, D., Mohlke, K. L., and Abecasis, G. R. (2008). Newly identified loci that influence lipid concentrations and risk of coronary artery disease. *Nature Genetics*, 40(2):161–169.
- [Wishart et al., 2012] Wishart, D. S., Jewison, T., Guo, A. C., Wilson, M., Knox, C., Liu, Y., Djoumbou, Y., Mandal, R., Aziat, F., Dong, E., Bouatra, S., Sinelnikov, I., Arndt, D., Xia, J., Liu, P., Yallou, F., Bjorn Dahl, T., Perez-Pineiro, R., Eisner, R., Allen, F., Neveu, V., Greiner, R., and Scalbert, A. (2012). HMDB 3.0: The Human Metabolome Database in 2013. *Nucleic Acids Research*, 41(D1):D801–D807.
- [Wray et al., 2003] Wray, G. A., Hahn, M. W., Abouheif, E., Balhoff, J. P., Pizer, M., Rockman, M. V., and Romano, L. A. (2003). The evolution of transcriptional regulation in eukaryotes. *Molecular Biology and Evolution*, 20(9):1377–1419.
- [Wu et al., 2016] Wu, Y., Yao, Y.-G., and Luo, X.-J. (2016). SZDB: A Database for Schizophrenia Genetic Research. *Schizophrenia Bulletin*, page sbw102.
- [Yamashita et al., 2014] Yamashita, A., Hayashi, Y., Nemoto-Sasaki, Y., Ito, M., Oka, S., Tanikawa, T., Waku, K., and Sugiura, T. (2014). Acyltransferases and transacylases that determine the fatty acid composition of glycerolipids and the metabolism of bioactive lipid mediators in mammalian cells and model organisms. *Progress in Lipid Research*, 53:18–81.
- [Yang et al., 2017] Yang, X., Sun, L., Zhao, A., Hu, X., Qing, Y., Jiang, J., Yang, C., Xu, T., Wang, P., Liu, J., Zhang, J., He, L., Jia, W., and Wan, C. (2017). Serum fatty acid patterns in patients with schizophrenia: a targeted metabolomics study. *Translational Psychiatry*, 7(7):e1176.
- [Yap et al., 2010] Yap, I. K. S., Anglely, M., Veselkov, K. A., Holmes, E., Lindon, J. C., and Nicholson, J. K. (2010). Urinary Metabolic Phenotyping Differentiates Children with Autism from Their Unaffected Siblings and Age-Matched Controls. *Journal of Proteome Research*, 9(6):2996–3004.
- [Yu et al., 2012] Yu, G., Wang, L.-G., Han, Y., and He, Q.-Y. (2012). clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters. *OMICS: A Journal of Integrative Biology*, 16(5):284–287.
- [Yu et al., 2018] Yu, Q., He, Z., Zubkov, D., Huang, S., Kurochkin, I., Yang, X., Halene, T., Willmitzer, L., Giavalisco, P., Akbarian, S., and Khaitovich, P. (2018). Lipidome

alterations in human prefrontal cortex during development, aging, and cognitive disorders. *Molecular Psychiatry*.

[Zoerner et al., 2011] Zoerner, A. A., Gutzki, F.-M., Batkai, S., May, M., Rakers, C., Engeli, S., Jordan, J., and Tsikas, D. (2011). Quantification of endocannabinoids in biological systems by chromatography and mass spectrometry: A comprehensive review from an analytical and biological perspective. *Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids*, 1811(11):706–723.