



Skolkovo Institute of Science and Technology

ROLE OF BREAST MILK LIPID COMPOSITION
IN POSTNATAL BRAIN DEVELOPMENT

Doctoral Thesis

by

ALEKSANDRA MITINA

DOCTORAL PROGRAM IN LIFE SCIENCES

Supervisor
Professor Philipp Khaitovich

Moscow - 2020

© Aleksandra Mitina 2020

I hereby declare that the work presented in this thesis was carried out by myself at Skolkovo Institute of Science and Technology, Moscow, except where due acknowledgement is made, and has not been submitted for any other degree.

Candidate (Aleksandra Mitina)
Supervisor (Prof. Philipp Khaitovich)

Abstract

Breast milk is an indispensable resource for nutrition and development of the growing organism during the first months of life. While the variety of the proteins and carbohydrates of breast milk are well described, lipids and lipid-soluble components, or the lipidome, are still poorly studied. Although it is mainly lipids that play a crucial role in composition and functioning of such structures as the nervous system and brain - their quantitative and qualitative presence defines characteristics of cell membrane and thus ability to conduct nervous impulse.

Considering the earlier evidence that fatty acids (FAs) can be directly incorporated into the brain phospholipids and can be actively transported through blood brain barrier (Wong, 2016), and that the main source of long-chain poly-unsaturated FAs in postnatal pups is breast milk, while taking into account the fact that the human brain undergoes extensive growth during the first year of life (Innis, 2007), when evolutionary human baby would be exclusively breastfed, we hypothesized that breast milk and brain compositions have co-evolved at least in terms of lipid or, more specifically, lipids' FA content. Previous studies demonstrated that the human brain lipidome composition is distinct compared to the brain lipidome of other species, including chimpanzees (Bozek, 2015), and is formed during the first year of life (Li, 2017). In this study we tested whether these human-specific features of the brain lipidome composition are reflected in the main source of the lipids that the young brain uses - breast milk.

Recent advances in mass-spectrometry facilitated high-throughput profiling of lipids and metabolites in different tissues and body fluids. With the use of modern mass-spectrometry instrumentation coupled to ultra-high pressure chromatography it is now possible to detect tens

of thousands of molecules of interest (Want, 2013; Dunn, 2011). Properly optimized chromatography methods allow to separate and detect a great number of lipids and fat soluble compounds to describe the total lipidome (Sarafian, 2014; Anwar, 2015; Vorkas, 2015).

In this research we are using liquid chromatography coupled to mass-spectrometry to study composition of intact lipidome of breast milk across seven mammalian species and show species-specific profiles at the level of TAGs. Analysis of the hydrolyzed milk samples revealed 81 FA with species-specific concentration levels and the chain length ranging from 4 to 30 carbons; in the brain we were able to detect 31 FA with species-specific concentration levels. For the 31 FA present both in milk and brain, intensities are highly correlated for the species studied. Moreover, the differences between the human milk and the macaque milk correlate with the differences between the human brain and macaque brain, with four long-chain unsaturated FAs that have the most impact. These results emphasize the connection between the breast milk lipid composition and the composition of the brain and suggest the role of human-specific breast milk fatty acids in the development of the unique cognitive features of the human brain.

Publications

1. **Mitina, A.**, Mazin, P., Vanyushkina, A. et al. Lipidome analysis of milk composition in humans, monkeys, bovids, and pigs. *BMC Evol Biol* 20, 70 (2020).
<https://doi.org/10.1186/s12862-020-01637-0>
2. Chugunova, A., Loseva, E., Mazin, P., **Mitina, A.**, Navalayeu, T., Bilan, D., Vishnyakova, P., Marey, M., Golovina, A., Serebryakova, M., Pletnev, P., Rubtsova, M., Mair, W., Vanyushkina, A., Khaitovich, P., Belousov, V., Vysokikh, M., Sergiev, P., & Dontsova, O. (2019). LINC00116 codes for a mitochondrial peptide linking respiration and lipid metabolism. *Proc Natl Acad Sci U S A*, 116(11), 4940–4945.
<https://doi.org/10.1073/pnas.1809105116>

Acknowledgements

I want to thank my supervisor Philipp Khaitovich and all the highly professional, passionate and motivated team members: Waltraud Mair, Pavel Mazin, Yun-Mei Wang, Olga Efimova, Ekaterina Khrameeva, Anna Vanyushkina, Ilia Kurochkin, Elena Stekolschikova, Anna Tkachev, Vita Stepanova, Alina Chernova and Nikolai Anikanov.

Table of Contents

Abstract	3
Publications	5
Acknowledgements	6
Table of Contents	7
List of Abbreviations	8
Chapter 1. Introduction	11
1.1 Evolution of lactation	11
1.2 Role of lipids in brain development	15
1.3 Mass-spectrometry of lipids	20
Chapter 2. Materials and Methods	25
2.1 Preparation of milk	25
2.2 Milk Hydrolysis	29
2.3 Preparation of brain	32
2.4 Brain Hydrolysis	33
Chapter 3. Results	36
3.1 Breast milk intact lipidome composition analysis in the test cohort	36
3.2 Breast milk lipidome analysis in the main cohort	45
3.3 Analysis of the brain lipidome FA composition	57
3.4 Comparison of the brain and milk lipidome FA composition	69
Chapter 4. Discussion	79
Chapter 5. Conclusions	85
Bibliography	87
Appendices	106

List of Abbreviations

AA - arachidonic acid

ABC - ATP-binding cassette

ABCG2 - ATP-binding cassette super-family G member 2

ALA - alpha-linolenic acid

ANOVA - analysis of variance

ASD - autism spectrum disorders

ATP - adenosine triphosphate

BBB - blood-brain barrier

BCE - Before the Common Era

BH - Benjamini-Hocheberg correction for multiple testing

BTA - Bos taurus autosome

CB - cerebellum

CGP - choline glycerophospholipids

DAG - diacylglycerol

DGAT1 - Diacylglycerol O-Acyltransferase 1

DHA - docosahexaenoic acid

EGP - ethanolamine glycerophospholipids

ELOVL - elongation of very long chain fatty acids

EQC - extraction quality control

FA - fatty acid

GP - glycerophospholipids

HPLC-MS - high-performance liquid chromatography coupled to mass-spectrometry

IGP - inositol glycerophospholipids

KO - knockout

LA - linoleic acid

LC-MS - liquid chromatography coupled to mass-spectrometry

LCFA - long-chain fatty acid

MAG - monoacylglycerol

MDS - multidimensional scaling

MS - mass spectrometry

MTBE - methyl tert-butyl ether

MUFA - monounsaturated fatty acid

mya - million years ago

NMR - nuclear magnetic resonance spectroscopy

OzID - ozone-induced dissociation

PA - phosphatidic acid

PC - phosphatidylcholine

PB - Paterno-Buchii

PCA - principal component analysis

PC(O-) - alkyl phosphatidylcholine

PC(P-) - plasmalogen phosphatidylcholine

PE(P-) - plasmalogen ethanolamine

PFC - prefrontal cortex

PI - phosphatidylinositol

PIP2 - phosphatidylinositol 4,5-bisphosphate

PIP3 - phosphatidylinositol 3,4,5-trisphosphate

PS - phosphatidylserine

PS(P-) - plasmalogen phosphatidylserine

PUFA - polyunsaturated fatty acid

QC - quality control

RT - retention time

SFA - saturated fatty acid

SGP - serine glycerophospholipids

SM - sphingomyelin

TAG - triacylglycerol

UPLC-MS - ultra-high-performance liquid chromatography coupled to mass-spectrometry

ULCFA - ultra long-chain fatty acid

UVPD - ultraviolet photodissociation

VLCFA - very long-chain fatty acid

Chapter 1. Introduction

1.1 Evolution of lactation

It is known that milk is the natural source of all necessary nutrition for a growing baby, providing sugars, fats and proteins in adequate proportion. Extensive research of the molecular composition of breast milk facilitated by the development of modern high-throughput techniques revealed that milk goes far beyond the definition of food. Breast milk contains bioactive molecules and immune factors (Demmelmair, 2017; Lee, 2018). It is not just food for a baby, it is a mixture of enzymes, hormones, vitamins, growth and immune factors that provide proper development of youth (Ballard, 2013; Bhinder, 2017; Hampel, 2017). It was even shown that certain oligosaccharides in breast milk can promote the growth of particular bifidobacteria in the infant intestine (Ward, 2006). Yet, how the lactation process appeared in the evolution of terrestrial animals and what was it like back then - is poorly understood, mainly due to the lack of artifacts. Still evolution of lactation and breast milk gains much interest from the researchers.

First evidence of adaptations that could lead to lactation arose more than 300 mya in the late Carboniferous period, when synapsids - ancestors of all living mammals - diverged from the sauropsids - ancestors of other two classes of amniotes: reptiles and birds. The main challenge for the terrestrial animals was to maintain their eggs in the airy conditions and lack of moisture. While sauropsids evolved the way of producing highly calcified eggs, synapsids had to choose another strategy to protect their parchment-shelled eggs from desiccation (Oftedal, 2002;

Oftedal, 2012). They were excreting moist from their skin - and this is believed to be the beginning of evolution of lactation (Oftedal, 2002).

Interestingly, the ability to pass necessary components to the offspring appeared several times in different animal species. This was partially necessitated by the lack of water, and partially by the remoteness of the nesting place from the feeding grounds. There are several avian species that produce so-called 'birds milk', or crop milk: pigeons and doves, flamingoes and Emperor penguins. Crop milk is a substance produced from the glands of the upper intestinal tract, which contains fats, proteins, minerals, and even antibodies (Baitechman, 2007), immunoglobulins (Goudswaard, 1979), hormones, and growth factors (Frelinger, 1971; Shetty, 1992). The fact that milk-like substances evolved independently on several occasions allows us to speculate about lactation in dinosaurs - since their nesting habits resume those in birds and could potentially co-evolve with feeding patterns (Else, 2013).

Lactation is a feature of all living mammals, with some differences between marsupials, monotremes and placental animals. For example, monotremes are the ones lacking nipples and their milk is poured down, which can be considered more similar to the original proto-milk in the way it is delivered to the offspring. Marsupial neonates are born very altricial, and milking is an essential step for their development. In placental animals, milking strategy and milk composition depends on the species lifestyle and the ability to accompany offspring after birth - if lactation period is short, then milk will be very dense, allowing to transfer all the necessary nutrition to the baby within this short period of time (Hinde, 2011). For example, Hooded seals, who give birth on ice floes, lactate for 4 days only, producing milk with more than 60% of fat (Oftedal, 1993; Oftedal, 1995) to help their pups maintain in the severe environment. Those species that have the

opportunity to accompany their youth longer after birth (such as bovids and primates), produce more diluted milk.

For domesticated species, artificial selection has had an effect on the lipid composition of milk. Changes in the genetics of dairy cattle are now occurring faster than at any time in history (VandeHaar, 2016). Cattle were domesticated from their ancestor - aurochs - in the Middle Euphrates River Valley about 9000 BCE, and came to Europe around 4000 BCE (Bollongino et al., 2012; Scheu et al., 2015). Modern dairy cattle are thinner and less muscular than their ancestral relatives, they also have greater milk producing capacity. In the past years, milk productivity of dairy cattle arose due to technological and reproductive advances, improvement of animal care and nutrition. Recent selection approaches, including the implementation of computational population genetics and development of breeding values, allows to manage economically important traits, such as milk yield, wiser and faster (Scheifers and Weigel, 2012).

A few domestication-related genes that affect milk composition have been identified in cattle. The *DGAT1* gene on *Bos taurus* autosome (BTA) 14 is associated with 30% of the variation in fat percentage in milk (Grisart, 2002). KO of a gene encoding a protein with acylCoA:diacylglycerol acyltransferase activity (DGAT1) was shown to completely inhibit lactation in the mouse (Smith et al. 2000). Another gene that affects milk composition is *ABCG2* (Cohen-Zinder, 2005), a member of the ATP binding cassette (ABC) superfamily (Ejendal and Hrycyna, 2002; Gottesman et al., 2002). Analysis of different stages of mammary development revealed that *ABCG2* was greatly induced during late pregnancy and especially during lactation (Jonker et al., 2005).

With respect to its macronutrient composition, milk of a particular species can be either high sugar, or high fat, or low sugar and low fat, but cannot be both high in sugar and fat. The reason lies in the milk synthesis in a mammary gland and its specialized secretory cells - lactocytes. Mature lactocytes produce lactose, a unique oligosaccharide found only in milk, which draws much water under osmotic pressure, diluting milk and thus lowering concentrations of fat and protein. Later on, the synthesis of lactose becomes downregulated in a mammary gland, during the process of involution, so milk becomes lower in sugar and thus in water, and higher in fat and protein.

In some species including primates, milk produced during the first several days of lactation (which is called colostrum) is quite different from the milk produced during the following lactation period - it contains very low concentrations of fat and sugar. Instead, it is rich in immunoglobulins and other peptides that accumulate in a mammary gland preceding the process of lactose synthesis.

Macronutrients in our diet fulfill two requirements: sugars and fats are the sources of energy, and proteins serve as building blocks. However, there can be exceptions. For example, obligate carnivores metabolize proteins for energy - it is known that the main source of glucose for cats is catabolism of amino acids (Eisert, 2011). At the same time, some lipids, particularly, long chain polyunsaturated FAs (LCPUFA), are not digested and can be directly incorporated into the cell membranes.

1.2 Role of lipids in brain development

Lipids constitute the majority of the dry mass of the brain and have been associated with proper functioning as well as pathological conditions of the brain (Wenk, 2005). Instead of being used for energy storage, brain lipids are incorporated into the cell membranes, particularly in neurons and astrocytes and in the myelin sheaths wrapped around axons. As the components of cell membranes, lipids are critically important for maintaining membrane fluidity and conducting nervous impulses, the functions highly critical in the brain (Innis, 2007; Mulder, 2018).

Brain lipids are presented by phospholipid and sphingolipid classes, as well as cholesterol and cholesterol metabolites (Sastry, 1985). About 4–5% of wet weight (4% in the gray matter and 7% in the white) is present by glycerophospholipids (GP). Ethanolamine glycerophospholipids (EGP) constitute the majority - about 36% - of the phospholipids in the brain, of which plasmalogen ethanolamine (PE(P-)) represents more than a half (O'Brien, 1965). GPs show high contents of the PUFAs (Panganamala, 1971). Out of choline glycerophospholipids (CGP) phosphatidylcholine (PC) comprises about 33%, with PC(16:0_18:1) being the most abundant species (Rouser, 1965). Plasmalogen PC (PC(P-)) and alkyl PC(O-) species account for about 2% of CGP. Serine glycerophospholipids (SGP) amount to about 17% of all phospholipids and are present in the form of phosphatidylserine (PS) or serine plasmalogen (PS(P-)) and contain mainly stearic (18:0), oleic (18:1n-9), and docosahexaenoic acid (DHA, 22:6n-3;4,7,10,13,16,19-22:6) (O'Brien, 1965; Panganamala, 1971). Inositol glycerophospholipids (IGP) account for about 3% of the total phospholipids, with the main species being phosphatidylinositol (PI) and triphosphoinositide (PIP3), and some

diphosphoinositide (PIP₂) with stearic (18:0) and arachidonic acid (AA, 20:4n-6;5,8,11,14-20:4) (Sastry, 1985; Panganamala, 1971; Rouser, 1965). Glycerophosphates and phosphatidic acid (PA) represent about 2% of total phospholipids. Diphosphatidylglycerol (cardiolipin) represents around 0.2% of phospholipids (Sastry, 1985; Guan, 1994). The main FA of the brain phospholipids are palmitic acid (16:0), palmitoleic acid (16:1), stearic (18:0), oleic (18:1n-9), and linoleic (LA,18:2n-6;9,12-18:2), with lower concentrations of alpha-linolenic (ALA,18:3n-3;9,12,15-18:3) and AA (Sastry, 1985; Guan, 1994).

Sphingolipids consist of sphingomyelin (SM) which is about 15% of total sphingolipids; cerebrosides - 16% of total lipids, mostly present as galactosylceramide; sulfatides - 6% of total lipids; and gangliosides with stearic acid (18:0) which is over 80% of the total ganglioside FA content (Sastry, 1985).

The two most abundant FAs in the brain are AA and DHA. These are found in high proportions in the cell membranes of neural tissue and retina, playing an important role in neurogenesis and neurotransmission. AA and DHA can be synthesized from the precursors - LA and ALA, which are considered essential and thus must be provided with food, through the processes of chain elongation and desaturation (Sprecher, 2000).

Specifically, ALA is desaturated to 18:4n-3 (6,9,12,15-18:4) by delta-6 desaturase, then undergoes chain elongation to form 20:4n-3 (8,11,14,17-20:4), and then desaturated to eicosapentaenoic acid (EPA, 20:5n-3;5,8,11,14,17-20:5) by delta-5 desaturase, which then can be chain elongate to docosapentaenoic acid (DPA,22:5n-3;7,10,13,16,19-22:5). While mammals encode delta-6 and delta-5 desaturase in their genomes, delta-4 desaturase required to synthesize

22:6n-3 directly from 22:5n-3 is lacking. Instead, 22:5n-3 is first elongated to 24:5n-3(9,12,15,18,21-24:5), then desaturated to 24:6n-3(6,9,12,15,18,21-24:6) by delta-6 desaturase, and finally undergoes beta-oxidation to remove 2 carbons and produce 22:6n-3. AA is synthesized from LA through the same pathway. As opposed to n-3DPA, n-6DPA is produced by chain elongation, desaturation and beta-oxidation of AA. It can be incorporated into lipids instead of DHA if the DHA supply is not sufficient (Galli, 1971).

Among neural cells (neurons, astrocytes, oligodendrocytes, microglia) ability to synthesize DHA from ALA was demonstrated only in astrocytes (Moore, 1991). However, stable isotope tracer studies show that in humans, less than 1% of dietary ALA is converted to DHA, and while this parameter increased in women as compared to men, and especially during pregnancy (Galli, 1971) - increased supplementation with ALA did not increase blood plasma DHA concentrations of pregnant women or of their newborn babies. This allows us to conclude that dietary supplementation of LCPUFA is a preferred source of DHA, at least in humans.

Supplementation of LCPUFA is of particular importance for the developing brain, since it undergoes extensive growth during the first year of life. Human baby's brain grows from approximately 350 grams at birth to the average of 660 grams by the age of 6 months, and then increases to the average of 925 grams at one year reaching approximately 70% of the adults brain weight (1200-1400 grams) (Dekaban, 1987; Tau, 2010). Developing human brain consumes about 50% of the metabolic energy at the resting state (Grande, 1979). This rapid growth of brain weight can be explained by an increase in grey matter as a result of extensive arborization rather than increase in white matter and myelination processes (Tau, 2010; Knickmeyer, 2008). Neural synapses proliferation increases after birth reaching peak of the rates of synapses formation by 8

months (Tau, 2010; Levitt, 2003), and these synapses are especially rich in LCPUFA and particularly DHA (Innis, 2007).

The first year of life - underlined by the rapid growth of neural tissue, extension of grey matter and formation of synaptic contacts - coincide with the period when the main food of the baby is breast milk. Accordingly, milk represents not only the source of macro- and micronutrients, immune factors and hormones, but also the main provider of the essential FAs for the developing brain. The question raises - does the breast milk lipid composition in different mammalian species reflect the differences in lipid composition of brains of these species? Or more accurately: if we see the evolution of brain lipid composition, can we trace the corresponding changes in milk composition, since milk is the main source of lipids in the brain? What were the recent evolutionary adaptations in the brain, and how much are they reflected in milk? What can we conclude from studying the evolution of milk composition, and what can we do to get all the benefits from breastfeeding, particularly for brain development?

It is more interesting to study the evolution of breast milk composition if we think about milk as a compromise between baby's requirements for growth, and mothers need to maintain her own health and reproductive ability. Breastfeeding appeared as a feature that gave our terrestrial predecessors the competitive advantage. Yet, both the mother and her offspring are interested in successful reproduction of their taxa. On one hand, breastfeeding has to offer maximum benefit to the offspring, providing nutrition for growth and development, immune and antibacterial factors for defense, necessary vitamins and microelements for proper functioning of immature systems. On the other hand, lactation should not harm or threaten the mother's or

child's health, resulting in lack of some elements needed for child development from the milk composition.

The well-known examples of milk composition sub-optimality with respect to baby nutrition are the low concentrations of vitamin D and iron. The reason why milk is poor in vitamin D is very simple - our predecessors could synthesize sufficient amounts of vitamin in their skin under ultraviolet rays due to extensive exposure to natural sunlight. There was no need to provide it with milk. Now, some human populations do not get sufficient sunlight exposure, the process accompanied by the growing incidences of rickets - a disease, caused by the insufficient concentrations of vitamin D in blood. There is now growing evidence that vitamin D participates not only in the calcification and mineralization of bones, but it also plays an important role in the development of the nervous system (Ali, 2019). There are about 400 genes that appear to have vitamin d binding receptors (VDBR) in their promoter regions, suggesting that this hormone may play a role in gene regulation (McCann, 2008). Vitamin D was shown to influence the formation of synaptic contacts (Ali, 2019). Vitamin D deficiency was also detected in kids with autism spectrum disorders (ASD), with severe symptoms disappearing in about 75% of patients treated with high doses of vitamin D (Saad, 2016).

As for the iron - this element is obligatory for the growth of bacteria, mostly pathogenic, that's why its concentrations in milk are very low. All the iron in milk comes bound to lactoferrin. The receptors for this protein are located in the neonate's intestine, so when lactoferrin gets to the intestine and binds to its receptors - iron is released and can be absorbed. Although in these two examples the reasons for deficiency in milk are different, the idea is the same - something that is unnecessary or harmful in the particular environment will not be

produced by the mammary gland. And whatever is produced - would be of tremendous importance.

Considering the earlier evidence that FAs can be directly incorporated into the brain phospholipids and can be actively transported through BBB (Silver, 2016), and that the main source of LCPUFA in postnatal pups is breast milk, while taking into account the fact that the human brain undergoes extensive growth during the first year of life (Innis, 2007), when evolutionary human baby would be exclusively breastfed, we hypothesized that breast milk and brain compositions have co-evolved at least in terms of lipid or, more specifically, lipids' FA content. Previous studies demonstrated that the human brain lipidome composition is distinct compared to the brain lipidome of other species, including chimpanzees (Bozek, 2015), and is formed during the first year of life (Li, 2017). Thus, it is reasonable to test whether these human-specific features of the brain lipidome composition are reflected in the main source of the lipids that the young brain uses - breast milk.

1.3 Mass-spectrometry of lipids

Two primary analytical techniques that are widely used for measuring the levels of polar and nonpolar metabolic compounds (polar metabolites and lipids) are nuclear magnetic resonance spectroscopy (NMR) and mass spectrometry (MS) (Nicholson, 2002; 2008). These methods are relevant to a wide variety of applications, from disease progression and biomarker discovery, therapeutic drug response and toxicity, and also in assessing metabolic effects of nutrition, exercise, development, and aging. Metabolomics provides us with a different kind of information compared to other “omics” techniques: while genomics can tell us about cell's

potential, and transcriptomics can tell us what is going to happen, metabolomics describes us the current condition of the cell, which is the closest biological proximity to the phenotype. Another advantage of metabolomics is its lower cost compared to genomics and transcriptomics - if the analytical platform itself is expensive, the per-sample costs are low. Thus, when considering the need to process large numbers of samples, metabolomics allows screening in a high-throughput manner and at a lower overall cost.

Separation techniques such as liquid chromatography that precede mass spectrometry analysis are commonly used to reduce the complexity of mass spectra and provide additional information for identification of the compounds. In particular, because of the complex composition of some biofluids and the highly variable ranges of metabolite concentrations, detection and measurements greatly benefit from multidimensional separation by high performance (HPLC-MS) and ultra-high-performance liquid chromatography coupled to MS (UPLC-MS). Due to diversity in physicochemical properties, polar metabolites and lipids are often analyzed separately by two different LC-MS methods.

Extensive application of mass-spectrometry to lipids became possible after the invention of soft ionization techniques: matrix-assisted laser desorption ionization, MALDI, and electrospray ionization, ESI (Tanaka, 1988; Yamashita, 1984). Since then, many methods for extraction, separation, and measurement of lipids in biofluids and tissues have been developed that facilitated analysis of many biologically important species (Wenk, 2010). A big issue of mass-spectrometry of lipids is their high chemical diversity (Fahy, 2005), making it difficult to measure the lipidome of a cell or a tissue in one experiment, with this experiment giving only preliminary, or untargeted information. In order to study the compounds of interest, one can

apply fragmentation methods, or tandem mass-spectrometry, to get the detailed information of the building blocks that constitute these lipids. Another method for identification of lipids based on their fragment ions, which is widely used is shotgun lipidomics. There are also other targeted approaches based on tandem mass-spectrometry available now (Wenk, 2005; Blanksby and Mitchell, 2010; Wenk, 2010).

The field of lipidomics is still developing, with many challenges remaining, especially regarding mass spectrometric feature annotation and experimental reproducibility. Annotation of lipid features is complicated due to the diverse composition and structural variability of the molecules. As opposed to proteins where the sequence can be directly reconstructed from the genome, lipid molecules vary in composition depending on many factors such as substrate availability, enzyme activity and relative proximity of the substrate from the enzyme. Another challenging task is assignment of the double bond configurations, which in addition to the conventional tandem mass-spectrometry requires implementation of more sophisticated techniques, such as ozone-induced dissociation (OzID) (Mitchell, 2009), the use of so-called Paterno-Buchii (PB) photochemical reaction as is (Ma, 2014) or combined by ultraviolet photodissociation (UVPD) (Wäldchen, 2017).

In most classification systems lipids are considered from the structural and biosynthetic perspective, with the widely accepted classification based on their chemically functional backbone, dividing them into five groups: polyketides, acylglycerols, sphingolipids, prenols, and saccharolipids. However, for simplification of the analysis, fatty acyls were separated from polyketides, glycerophospholipids from glycerolipids, and sterol lipids from prenols, resulting in eight categories. This system employed by LIPIDMAPS Lipidomics Gateway and LIPID MAPS

Structure Database (LMSD) allows for subdivision of the main categories into classes and subclasses to handle the existing and emerging arrays of lipid structures (Fahy, 2005).

For example, abbreviations for glycerolipids (MAG,DAG,TAG) are used to refer to species with one (Mono-acyl), two (Di-acyl) or three (Tri-acyl) side-chains, respectively, where the structures of the side chains are indicated within parentheses indicating total chain length and number of double bonds (i.e., TAG (48:2)). FAs can be classified according to their carbon (C) chain-length, and the number of double bonds, where long-chain FAs (LCFAs) have chain-lengths of C11–20; FAs longer than C20 (C>20) are called very long-chain FAs (VLCFAs). Another classification of FAs is based on the number of double bonds. FAs are classified into saturated FAs (SFAs) without double bonds; monounsaturated FAs (MUFAs) with one double bond; and polyunsaturated FAs (PUFAs) with two or more double bonds. PUFAs are further sub-classified into n-3 (or ω 3) and n-6 (or ω 6) series depending on the position of the double bond most distant from the carboxyl group (Sassa, 2014).

Lipids are commonly defined as biological substances that are hydrophobic in nature and soluble in organic solvents (Smith, 2000). A more accurate definition of a lipid would be a hydrophobic or amphipathic compound that may originate entirely or in part by carbanion-based condensations of isoprene units (prenols, sterols, and others). Lipids are also subdivided into ‘simple’ and ‘complex’ ones, with simple lipids yielding two types of products of hydrolysis (i.e., FAs, sterols, and acylglycerols), and complex lipids yielding three or more products of hydrolysis (i.e., glycerophospholipids and glycosphingolipids). Any classification system is

subjective as a result of structural and biosynthetic complexity of lipids but is essential for appropriate research and data management.

Unification of research in terms of data acquisition and downstream analysis is another big challenge for the lipidomics community. As the interlaboratory comparison exercise indicated (Bowden, 2017), measurements of the same mixture of standard blood plasma (1950 metabolites) conducted in 23 different lipidomic laboratories, showed high variability in the reported results. This study raised the need to unify the protocols of samples handling, extraction, and spectra acquisition and led to the establishment of Lipidomics Standards Initiative Consortium that was launched in spring 2018 to address these challenges (Consortium LS, 2019).

Chapter 2. Materials and Methods

2.1 Preparation of milk

Human breast milk collection was organized in two locations: in Moscow from June, 2015 - through November, 2015, and in Shanghai from May, 2015 - through September, 2016. There were 91 participants from Moscow, and 87 from Shanghai, all of them participated voluntarily, have familiarized with the study objectives and signed Informed Consent Form. Each of them was asked to collect milk at the end of feeding in order to obtain 'back' milk, which comes at the end of breastfeeding and is more rich in fat, as opposed to 'fore' milk, which comes at the beginning of feeding and is more rich in carbohydrates and water. All participants received the same set of jars, where they were collecting milk once a week, and stored it at -20 °C. Every week these milk samples were transferred to the lab where they were stored at -80 °C prior to analysis. In total we collected 297 samples from Moscow, and 291 samples from Shanghai.

To make aliquots, frozen milk samples were cut into vertical slices to ensure all layers of milk are equally present in an aliquot and to avoid uneven fat content due to delamination. These aliquots were thawed until 0 °C and 16 µl of milk were transferred to a 2.0 ml Eppendorf safe-lock tube and resuspended in 34 µl of LC-MS grade water. For the test cohort, samples were randomized by species prior to extraction, extraction was performed in one round within one day. For the main cohort, milk samples were extracted in batches with 72-96 samples in each batch within a 7-day period. In order to control for the batch-effect introduced by extraction, we made a mix sample and extracted it every 24-th sample, providing us with extraction quality

control (EQC). Separately we performed the same extraction procedure for 24 probes (extraction blanks) containing only buffers to control for contaminants that might appear during extraction and that do not originate from milk.

For total lipid extraction, modified 2-phase protocol was used (Sarafian, 2014), all manipulations with the samples were performed on ice:

- 750 ul of MeOH:MTBE (1:3) solution containing internal standards in concentration of 1 mg/L were added to each sample, vortexed for 1 min, sonicated for 15 min in an ice-cooled sonication bath, incubated for 30 min at 4 °C, and sonicated for the second time in a pre-cooled sonication bath
- 560 ul of MeOH:H₂O (1:3) solution was then added to each sample, vortexed for 10 sec and centrifuged for 10 min at 14.000 x g at 4 °C
- 400 ul of the upper-phase, containing organic fraction, was transferred to a new 2.0 ml Eppendorf tube and dried in a speedvac for 1 h at 30 °C.

Dried lipids were then resuspended in 400 ul of acetonitrile:isopropanol (1:3) solution, samples were rigorously vortexed for 10 sec, shaken for 10 min at 4 °C and sonicated in an ice bath. 5 ul of each sample were transferred to the autosampler glass vial and diluted 1:20 with 95 ul of acetonitrile:isopropanol (1:3) solution. For test cohort, pool of all samples to perform quality control (QC) was prepared by mixing 5 ul from each sample in a separate Eppendorf tube, transferred to glass vials and diluted 1:20 with acetonitrile:isopropanol (1:3) solution to get quality control samples. Pools for each species were made by mixing 10 ul of the samples of the corresponding species and diluted 1:20 with acetonitrile:isopropanol (1:3) solution. For the main

cohort, QC samples were prepared by mixing all samples left from the test cohort. 3 µl per diluted sample were separated on a Reversed Phase Bridged Ethyl Hybrid (BEH) C8 reverse column (100 mm x 2.1 mm, containing 1.7 µm diameter particles, Waters) coupled to a Vanguard pre-column with the same dimensions, using a Waters Acquity UPLC system (Waters, Manchester, UK). The mobile phases used for the chromatographic separation were: water, containing 10 mM ammonium acetate, 0.1% formic acid (Buffer A) and acetonitrile:isopropanol (7:3 (v:v)), containing 10 mM ammonium acetate, 0.1 % formic acid (Buffer B). The gradient separation was: 1 min 55 % B, 3 min linear gradient from 55 % to 80 % B, 8 min linear gradient from 80 % B to 85 % B, and 3 min linear gradient from 85 % A to 100 % A. After 4.5 min washing with 100 % B the column was re-equilibrated with 55 % B. The flow rate was set to 400 µl/min. The column temperature was maintained at 60 °C.

For test cohort mass spectra of the total milk lipidome were acquired in positive mode using a heated electrospray ionization source in combination with a Bruker Impact II QTOF (quadrupole-Time-of-Flight) mass spectrometer (Bruker Daltonics, Bremen, Germany) and analyzed separately. 4 blank samples were run at the beginning of the queue, followed by 4 QC samples to equilibrate the column. 38 samples were queued in the same order as for extraction (with all samples randomized by species), interleaved with 7 pools, one QC preceding the first sample and then a QCs after every 9th sample. At the end of the queue, 2 injections containing 100% acetonitrile ran to wash the column, followed by blank samples - prepared as a usual sample but containing only extraction buffers to reveal all contaminants that could come from extraction but not from the sample.

For the main cohort mass spectra of the total milk lipidome were acquired in positive mode using a heated electrospray ionization source in combination with QExactive Hybrid Quadrupole-Orbitrap mass spectrometer (Thermo Scientific, Germany). MS settings in positive acquisition mode were as follows: spray voltage was set to 4.5 kV, S-lens RF level at 70, and heated capillary at 250 °C; aux gas heater temperature was set at 350 °C; sheath gas flow rate was set to 45 arbitrary units; aux gas flow rate was set to 10 arbitrary units; sweep gas flow rate was set to 4 arbitrary units. Full scan resolutions were set to 70 000 at m/z 200. Full scan target was 10e6 with a maximum fill time of 50 ms. The spectra were recorded using full scan mode, covering a mass range from 100–1500 m/z.

All samples (n = 859) were divided into 94-sample batches, which resulted in 9 complete batches and one short batch. All species were equally represented in each batch; all samples in a batch were randomized by species; human samples were also stratified by population, sex, age, parity and the mode of delivery. Final MS-queue was preceded by 8 blank injections (acetonitrile:isopropanol solution) and 6 QC samples to equilibrate the column. QC sample was injected after every 12-th milk sample, EQC was injected after every 24-th sample; the queue was followed by 5 blank injections and 8 washes. At the end of the queue, 24 extraction blanks were injected followed by 2 washes.

In total we acquired 1040 files for the main cohort, that were converted to .mzML format using ProteoWizard, Filters: Peak Picking: Vendor msLevel = 1; Subset: scan Time (360, 1050). Samples were then arranged into folders according to the batches: Batch_0 - Batch_8, plus batch 'washes-banks'. To find the optimal parameters for the dataset, IPO (Libiseller, 2015) was run on 9 QC files, one from each batch (QC-7, 10, 16, 25, 33, 42, 51, 60, 69). The best parameters

were as follows: peakwidth = c(8, 36), ppm = 12.5; noise level was set to 10^6 , snthresh set to 100, minfrac set to 0.1. Xcms (Smith, 2006) was run on all files organized by batches, alignment performed with the center sample MS14 (crab-eating macaque), followed by CAMERA (Kuhl, 2012). From the output table, containing 2848 lipid features, we kept only monoisotopic peaks, resulting in 1856 features. We excluded peaks with average signal intensity in blanks higher than 10% of the average intensity in milk samples, and peaks with coefficient of variation (CoV, calculated as standard deviation over the mean) higher than 30% in the QC samples. This resulted in 647 lipid features that were subject for further statistical analysis. Remaining 647 lipid features were upper-quartile normalized and log-transformed.

2.2 Milk Hydrolysis

In order to assess the FA composition that comprises milk lipids, we performed hydrolysis (Bromke, 2015). 100 ul of milk lipid extracts, prepared previously for total lipidome measurements, were transferred to a separate Eppendorf tube and dried in speedvac at 40 °C until dry. Samples were hydrolysed in batches of 96 samples each within a 7-day period. 12 probes (extraction blanks) containing only buffers were extracted with the same procedure to control for contamination.

The procedure was as follows:

- 100 ul of methanol : 6% KOH (4 : 1) solution containing FA internal standards was added to dried samples and incubated for 2 h at 60° with continuous shaking
- 100 ul of saturated NaCl solution was added after cooling to room temperature
- 50 ul of 29% HCl solution was added to each sample followed by 1 min vortexing

- 200 ul of CHCl₃:Heptane (1:4) solution was added to each sample, vortexed for 10 min and centrifuged for 5 min at 4 °C; upper phase was collected to the new Eppendorf tube; the rest of the solution was washed with 200 ul of CHCl₃:Heptane (1:4) and vortexed for 1 min; upper phase was collected
- 200 ul of UPLC H₂O was added to the Eppendorf tube containing FA extract, vortexed for 1 min; upper phase was collected to the new Eppendorf tube and dried in a speedvac for 1 hour.

Dried lipids were resuspended in 100 ul of acetonitrile:isopropanol (1:3) solution, samples were rigorously vortexed for 10 sec, shaken for 10 min at 4 °C and sonicated in an ice bath. 50 ul of each sample were transferred to the autosampler glass vial. Pool of all samples to perform quality control (QC) was prepared by mixing 10 ul from each sample in a separate Eppendorf tube, and transferred to glass vials to get quality control samples. 3 µl of prepared sample were separated on a Reversed Phase 1.8 µm High Strength Silica (HSS) T3 reverse column (100 mm x 2.1 mm, containing 1.8 µm diameter particles, Waters) coupled to a Vanguard pre-column with the same dimensions, using a Waters Acquity UPLC system (Waters, Manchester, UK). The mobile phases used for the chromatographic separation were: water, containing 10 mM ammonium acetate, 0.1% formic acid (Buffer A) and acetonitrile:isopropanol (7:3 (v:v)), containing 10 mM ammonium acetate, 0.1 % formic acid (Buffer B). The gradient separation was: 1 min 55 % B, 3 min linear gradient from 55 % to 80 % B, 8 min linear gradient from 80 % B to 85 % B, and 3 min linear gradient from 85 % A to 100 % A. After 4.5 min washing with 100 % B the column was re-equilibrated with 55 % B. The flow rate was set to 400 µl/min. The column temperature was maintained at 40 °C.

Mass spectra were acquired in negative ionization mode using a heated electrospray ionization source in combination with QExactive Hybrid Quadrupole-Orbitrap mass spectrometer (Thermo Scientific, Germany). MS settings in the negative ionization mode were as follows: spray voltage was set to 3 kV, S-lens RF level at 70, and heated capillary at 250 °C; aux gas heater temperature was set at 350 °C; sheath gas flow rate was set to 45 arbitrary units; aux gas flow rate was set to 10 arbitrary units; sweep gas flow rate was set to 4 arbitrary units. Full scan resolutions were set to 70 000 at m/z 200. Full scan target was 10e6 with a maximum fill time of 50 ms. The spectra were recorded using full scan mode, covering a mass range from 100– 1500 m/z. All samples were divided into 94-sample batches, randomized and ordered in the same way as for total lipidome measured in positive ionization mode.

For milk hydrolyzed data, all samples were converted to .mzML format using ProteoWizard, Filters: Peak Picking: Vendor msLevel = 1; using the entire time range. Optimized parameters for peak picking and alignment were as follows: peakwidth = c(7.5, 34), ppm = 12; noise level was set to 10^6 , snthresh set to 100, minfrac set to 0.1. Samples were arranged into folders according to the batches. Xcms was run on all files (921 in total) organized by batches, alignment performed with the center sample MS522 (crab-eating macaque), followed by CAMERA. The output table contained 1617 peaks, with remaining 1250 monoisotopic, which were then subject to filtering. Out of all peaks, we filtered out those with intensity in blanks of more than 10% of the intensity in real samples (considering they represent contamination). From those, only peaks with the coefficient of variation lower than 30% in all QC samples were kept, as they were reproducible. This resulted in 516 peaks.

For the annotation of hydrolyzed milk dataset we generated theoretical lists with all possible masses of the FAs with carbon chain length from 10 to 30, and double bond number from 0 to 6, considering also odd-chain FAs, as they were present in milk. 81 features from detected lipid peaks matched theoretical masses in our target list with the accuracy within 10 ppm and showed good agreement with mz-RT plot (*fig. 7A*).

2.3 Preparation of brain

Human brain samples (n = 92) were obtained from National Institute of Child Health and Human Development (NICHD), representing african american, caucasian, chinese and hispanic populations. We considered two brain regions: prefrontal cortex (PFC) and cerebellum (CB). All these samples were taken postmortem from children from newborns until under 1 year; meta information included age in days and sex. Chimpanzee brain samples (n = 18), macaques brain samples (n = 38), goat and pig brains samples (n = 26 and n = 20) were representing two brain regions - prefrontal cortex and cerebellum; meta information included age in days and sex (supplementary *tbl.3*). Prior to extraction, brain samples were cut into 10-15 mg pieces on ice to avoid melting, and placed into 2 ml reinforced Precellys tubes containing 2.8 mm zirconia beads.

All brain samples (n = 194) were stratified by species, region, age, sex and extracted in two batches. Extraction was performed by two-phase method (Sarafian, 2014) preceded by homogenization:

- 500 ul of MeOH:MTBE (1:3) solution containing internal standards in concentration of 1 mg/L was added to each sample

- homogenization was performed with the following program: 30 sec intensive stirring at 5000 rpm, 30 sec cooling down; 4 cycles
- sonication for 15 min in an ice-cooled sonication bath, incubation for 30 min at 4 °C, sonication for the second time in a pre-cooled bath
- 700 ul of MeOH:H₂O (1:3) solution was then added to each sample, vortexed for 10 sec and centrifuged for 10 min at 14.000 x g at 4 °C
- 400 ul of the upper-phase, containing organic fraction, was transferred to a new 2.0 ml Eppendorf tube and dried in a speedvac for 1 h at 30 °C.

2.4 Brain Hydrolysis

Hydrolysis of brain samples was performed to assess the FA composition of brain lipids (Bromke, 2015). Samples were hydrolyzed in two batches; 12 probes (extraction blanks) containing only buffers were extracted to control for contamination.

The procedure was as follows:

- 100 ul of methanol : 6% KOH (4 : 1) solution containing FA internal standards was added to dried samples and incubated for 2 h at 60° with continuous shaking
- 100 ul of saturated NaCl solution was added after cooling to room temperature
- 50 ul of 29% HCl solution was added to each sample followed by 1 min vortexing
- 200 ul of CHCl₃:Heptane (1:4) solution was added to each sample, vortexed for 10 min and centrifuged for 5 min at 4 °C; upper phase was collected to the new Eppendorf tube; the rest of the solution was washed with 200 ul of CHCl₃:Heptane (1:4) and vortexed for 1 min; upper phase was collected

- 200 μ l of UPLC H₂O was added to the Eppendorf tube containing FA extract, vortexed for 1 min; upper phase was collected to the new Eppendorf tube and dried in a speedvac for 1 hour.

Dried lipids were resuspended in 100 μ l of acetonitrile:isopropanol (1:3) solution, samples were rigorously vortexed for 10 sec, shaken for 10 min at 4 °C and sonicated in an ice bath. 50 μ l of each sample were transferred to the autosampler glass vial. Pool of all samples to perform quality control (QC) was prepared by mixing 10 μ l from each sample in a separate Eppendorf tube, and transferred to glass vials to get quality control samples. 3 μ l of prepared sample were separated on a Reversed Phase 1.8 μ m High Strength Silica (HSS) T3 reverse column (100 mm x 2.1 mm, containing 1.8 μ m diameter particles, Waters) coupled to a Vanguard pre-column with the same dimensions, using a Waters Acquity UPLC system (Waters, Manchester, UK). The mobile phases used for the chromatographic separation were: water, containing 10 mM ammonium acetate, 0.1% formic acid (Buffer A) and acetonitrile:isopropanol (7:3 (v:v)), containing 10 mM ammonium acetate, 0.1 % formic acid (Buffer B). The gradient separation was: 1 min 55 % B, 3 min linear gradient from 55 % to 80 % B, 8 min linear gradient from 80 % B to 85 % B, and 3 min linear gradient from 85 % A to 100 % A. After 4.5 min washing with 100 % B the column was re-equilibrated with 55 % B. The flow rate was set to 400 μ l/min. The column temperature was maintained at 40 °C.

Mass spectra were acquired in negative ionization mode using a heated electrospray ionization source in combination with QExactive Hybrid Quadrupole-Orbitrap mass spectrometer (Thermo Scientific, Germany). MS settings in the negative ionization mode were as follows: spray voltage was set to 3 kV, S-lens RF level at 70, and heated capillary at 250 °C; aux

gas heater temperature was set at 350 °C; sheath gas flow rate was set to 45 arbitrary units; aux gas flow rate was set to 10 arbitrary units; sweep gas flow rate was set to 4 arbitrary units. Full scan resolutions were set to 70 000 at m/z 200. Full scan target was 10e6 with a maximum fill time of 50 ms. The spectra were recorded using full scan mode, covering a mass range from 100– 1500 m/z. Samples were divided into two batches, randomized and ordered in the same way as for total lipidome measured in positive ionization mode.

For brain hydrolyzed dataset, all samples were converted to .mzML format using ProteoWizard, Filters: Peak Picking: Vendor msLevel = 1; entire time range. Samples were then arranged into folders according to the batches. To find the optimal parameters for the dataset, IPO was run on 3 QC files, one from each batch (QC-5, 10, 17). The best parameters were as follows: peakwidth = c(7.5, 34), ppm = 12; noise level was set to 10^6 , snthresh set to 100, minfrac set to 0.1. Xcms was run on all files (290 in total) organized by batches, alignment performed with the center sample MS200 (human cerebellum), followed by CAMERA. The output table, after removing isotopes, contained 211 peaks, which were then subject to filtering. Out of all peaks, we filtered out those with intensity in blanks of more than 10% of the intensity in real samples (considering they represent contamination). From those, only peaks with the coefficient of variation lower than 30% in all QC samples were kept.

For the annotation of hydrolyzed brain dataset we generated theoretical lists with all possible masses of the FAs with carbon chain length from 10 to 30, and double bond number from 0 to 6, considering also odd-chain FAs, as they were present in milk. 31 features from detected lipid peaks matched theoretical masses in our target list and showed good agreement with m/z-RT plot (*fig. 12A* and *fig. 12D*).

Chapter 3. Results

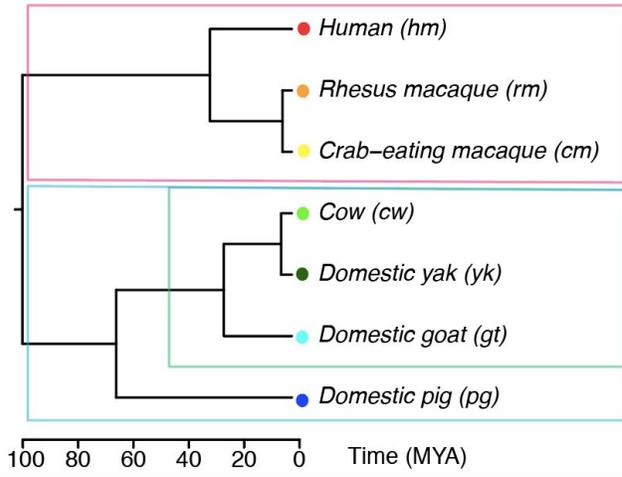
3.1 Breast milk intact lipidome composition analysis in the test cohort

Analysis of milk samples from the test cohort reveals differences in lipid milk composition across species. The cohort included breast milk samples from healthy human volunteers representing Russian (Moscow) and Chinese (Shanghai) populations (n = 20), cows (n = 4), goats (n = 4), pigs (n = 4), domestic yaks (n = 2), rhesus monkeys (n = 2), and crab-eating monkeys (n = 2). Furthermore, for each species we measured a pooled sample containing equal volumes of milk from each individual.

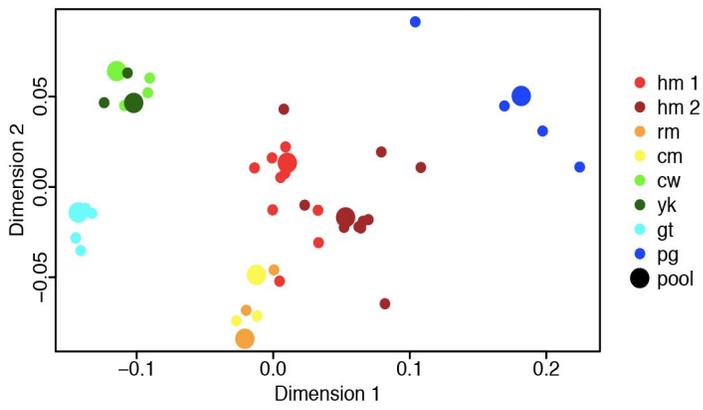
Phylogenetically, examined species represent two orders of mammals: artiodactyla, or cloven-hoofed mammals, and primate. Within artiodactyla, our study contains representatives of two families: suidae (pig) and bovidae (cow, domestic yak, and goat). Primate species were represented by cercopithecidae (crab-eating and rhesus monkeys) and hominidae (human) (*fig. 1A*).

Milk samples were extracted and measured in random order. The measurements were carried out using untargeted mass spectrometry coupled with high performance liquid chromatography in positive ionization mode. The measurements yielded a total of 472 mass spectrometry features representing distinct hydrophobic compounds (lipids) with molecular weights below 1,200 Daltons (Da).

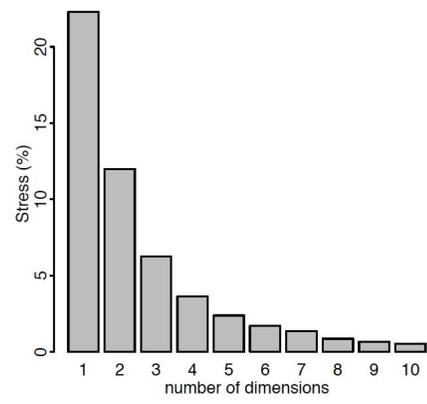
A



B



C



D

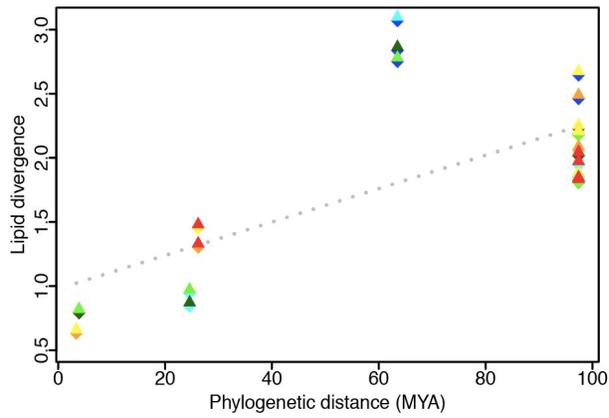


Fig. 1. Phylogenetic relationship among studied species. (A) Phylogenetic tree of 7 mammalian species. Colored dots indicate species. Colored frames indicate three groups: primates, suidae and bovidae. (B) Relationship between the samples based on signal intensities of 472 detected lipid features plotted in two dimensions using multidimensional scaling algorithm. Colors represent species: human from Moscow (hm1) - red, human from Shanghai (hm2) - brown, rhesus macaque (rm) - orange, crab-eating macaque (cm) - yellow, cow (cw) - green, domestic yak (yk) - dark green, domestic goat (gt) - cyan, domestic pig (pg) - indigo; each dot represents one sample; large dots represent pools of samples from the same species. The residual variance calculated for the MDS is 12%. (C) Value of the stress by the number of dimensions used in the MDS. (D) Relationship between species-specific lipid concentration divergence and phylogenetic distance for each pair of species. Each rhombus represents a pair of species; colors indicate species.

Visualization of the relationship among samples based on the abundance levels all 472 detected lipids using multidimensional scaling (MDS) revealed good separation of species and phylogenetic groups (*fig. 1B*). Furthermore, distances between species calculated using the normalized intensities of mass spectrometry signals, generally agreed with the phylogenetic distances (*fig. 1C*). Notably, lipid composition of the pig milk stood out from this linear relationship, showing greater difference to the *bovidae* species than expected from the phylogeny.

Statistical analysis of the lipid intensity differences identified 403 (85%) of the 472 detected lipids as significantly diverged among species (ANOVA, BH-corrected $P < 0.05$). Unsupervised clustering of these 403 lipids based on their intensity profiles across samples yielded 4 clusters (*fig. 2A*). Notably, lipid intensities within these clusters not only differed between phylogenetic families - with lipids higher abundant in pigs as compared to bovids and

primates (cluster 1), lipids higher abundant in bovids as compared to pigs and primates (cluster 3), and lipids higher abundant in primates and pigs as compared to bovids (cluster 2), but also within the families - with lipids higher abundant in macaques as compared to humans and pigs as compared to bovids (cluster 4) (*fig. 2B*).

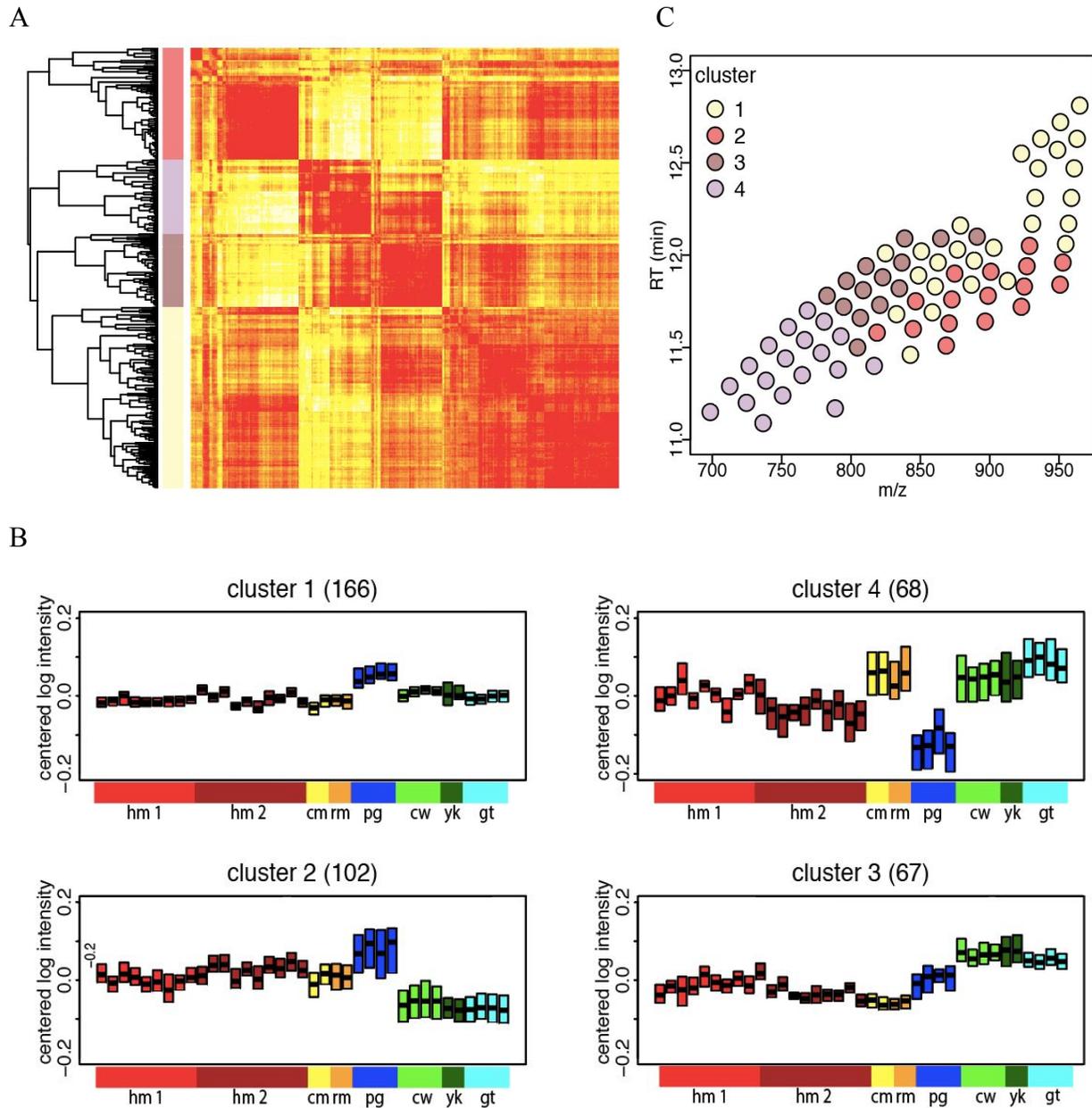
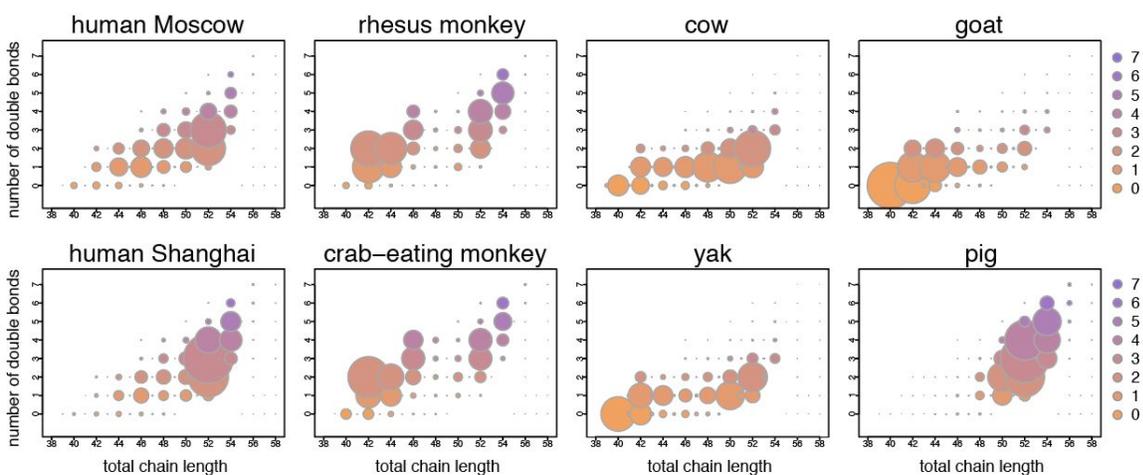


Fig. 2. Clusterization analysis of species-specific features. (A) Correlation matrix of species-specific lipids based on (1-cor) Pearson's distances. Clusterization performed by hierarchical clustering with complete linkage, left panel coloured by clusters: pink - 2nd cluster, grey - 4th cluster, brown - 3rd cluster and yellow - 1st cluster. (B) Distribution of centered signal intensities of lipids for a given cluster. Each box represents one sample. Number of features is indicated at the top. Samples coloured and ordered by species. (C) 76 annotated $[M + NH_4]^+$ TAG peaks are plotted coloured by clusters. The x-axis represents the mass of the compound, m/z . The y-axis represents retention time of the compound in minutes, RT (min).

In agreement with previous knowledge, milk samples showed prevalence of glycerolipids, and more specifically, triacylglycerides (TAGs), cumulatively constituting 98% of the total lipids (Wastra, 1999). Interestingly, annotated TAGs were present in all four clusters, covering the entire spectrum of lipid variation patterns: very long-chain (cluster 1), long-chain poly-unsaturated (cluster 2), long-chain (cluster 3), and middle-chain (cluster 4) species (*fig. 2C*). To investigate this further, we focused on differences in the abundance of 76 TAGs detected across species and having even cumulative FA chain length.

Remarkably, relative abundance analysis of these TAGs revealed clear intensity patterns characteristic of each mammalian species. Specifically, cow's milk contains more TAGs composed by long-chain monounsaturated fatty acids. By contrast, goat milk contains more TAGs composed of short and medium-chain saturated and monounsaturated fatty acids. Pigs milk stands from the rest of the species by having TAGs composed of long-chain polyunsaturated FAs. Monkey milk has more TAGs composed of medium-chain monounsaturated and polyunsaturated FAs. Finally, human milk tends to contain more TAGs with medium-chain polyunsaturated FAs (*fig. 3A*).

A



B

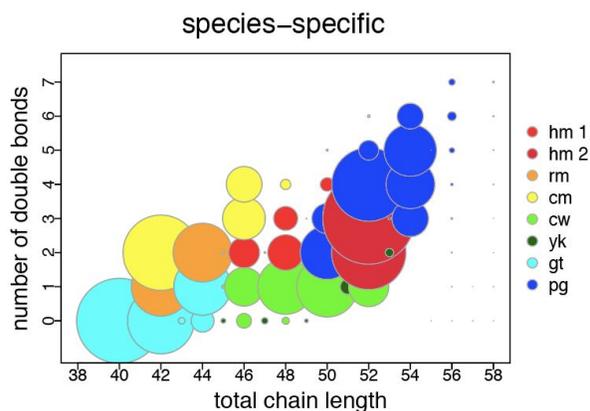
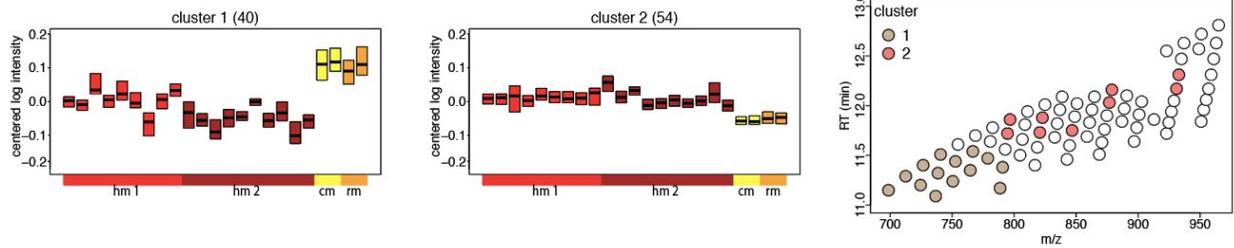


Fig. 3. Species-specific TAG profiles. (A) Distribution of TAGs from breast milk for each animal species. The x-axis represents the total chain length. The y-axis represents the number of double bonds; each point is a separate TAG. Intensity of each TAG is calculated as the mean intensity of the TAG in this species divided by the sum of mean intensities of all TAGs in this species. Size of each dot represents the calculated intensity of TAG multiplied by 0.3 and plus 0.5. Color of each dot represents the number of double bonds with gradients from orange - zero double bonds, to purple - seven double bonds. (B) Species-specific distribution of TAGs from breast milk for all animal species together. Y-axis represents the number of double bonds, x-axis represents the total chain length; each point is a separate TAG. Intensity of each TAG is calculated as: maximum intensity calculated for this TAG across all species. Size of each dot

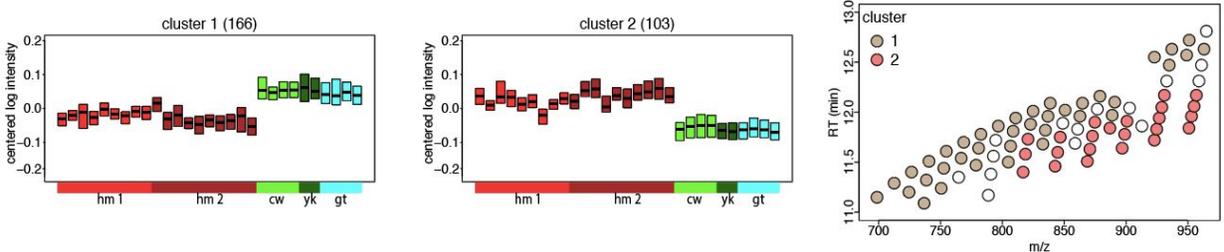
represents the calculated intensity of TAG multiplied by 0.3 and plus 0.5. Color of each dot represents the species with maximum calculated intensity for this TAG.

We next specifically searched for lipids showing statistically significant abundance differences between humans and all other three families represented in our study. Of the 472 detected lipids, 94 showed such differences between humans and macaques, 23 of them annotated TAGs (ANOVA, BH-corrected $P < 0.05$). Clustering of these lipids further revealed a specific pattern containing 54 lipids, with intensities specifically elevated in human milk. Notably, 9 TAGs constituting this pattern all contained medium and long-chain saturated, mono- and polyunsaturated FAs (*fig. 4A*).

A



B



C

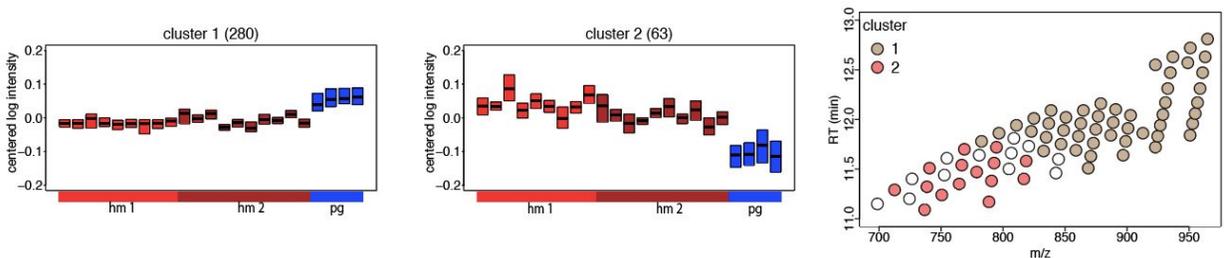


Fig. 4. Inter-species comparison of TAG profiles. (A) Distribution of signal intensities of TAGs between humans and macaques. Each box represents one sample. Samples are ordered and coloured by species. Number of TAGs is indicated on top of the box plot. 76 annotated (M + NH₄)⁺ TAG peaks are plotted coloured by clusters. Far right plot shows 76 annotated TAG species in mass over charge (m/z) by retention time (RT) coordinates. All circles represent annotated TAGs. Colored are TAGs with significantly different intensity levels: beige if higher in macaques (cluster 1), red if higher in humans (cluster 2). (B) Distribution of signal intensities of TAGs between humans and bovids. Each box represents one sample. Samples are ordered and coloured by species. Number of TAGs is indicated on top of the box plot. 76 annotated (M + NH₄)⁺ TAG peaks are plotted coloured by clusters. Far right plot shows 76 annotated TAG species in mass over charge (m/z) by retention time (RT) coordinates. Colored are TAGs with significantly different intensity levels: beige if higher in bovids (cluster 1), red if higher in humans (cluster 2). (C) Distribution of signal intensities of TAGs between humans and pigs. Each box represents one sample. Samples are ordered and coloured by species. Number of TAGs is indicated on top of the box plot. 76 annotated (M + NH₄)⁺ TAG peaks are plotted coloured by clusters. Far right plot shows 76 annotated TAG species in mass over charge (m/z) by retention time (RT) coordinates. All circles represent annotated TAGs. Colored are TAGs with significantly different intensity levels: beige if higher in pigs (cluster 1), red if higher in humans (cluster 2).

Comparison between humans and bovids revealed 269 lipids, 61 of them annotated as TAGs (ANOVA, BH-corrected $P < 0.05$). There were 103 lipids, with intensities specifically elevated in humans, 23 TAGs constituting this pattern contained medium and long chain polyunsaturated FAs (*fig. 5B*). Comparison between humans and pigs revealed 343 lipids, 64 of them annotated as TAGs (ANOVA, BH-corrected $P < 0.05$). There were 63 lipids, with intensities specifically elevated in humans, interestingly 15 TAGs constituting this pattern

contained medium chain saturated, mono- and polyunsaturated FAs, as opposed to long-chain polyunsaturated TAGs, elevated in pigs (*fig. 4C*).

3.2 Breast milk lipidome analysis in the main cohort

The main cohort included 297 human milk samples collected in Moscow and 291 human milk samples collected in Shanghai; animal samples included 66 cow and 35 goat samples collected in Moscow; 39 cow, 49 pig, 15 yak, 22 crab-eating and 23 rhesus macaque milk samples collected in Shanghai. All women have signed an Informed Consent Form. For human samples we collected information that was reported to affect breast milk composition (Hinde 2009, Powe 2010, Hinde 2013, Thakkar 2013, Galante 2018) such as: lactation stage (number of days from the childbirth), parity (number of children a woman had previously had), mode of delivery (natural birth or Caesar-section) and baby's sex (supplementary *tbl.1*). We also collected information on mother's age, weight and height, baby's weight and height at birth (supplementary *tbl.2*).

Analysis of the intact milk lipidome of the main cohort confirmed separation of species and phylogenetic groups observed in the test. Representatives of *bovidae* family (cow, domestic yak and goat) from *artiodactyla* grouped together, while representatives of *suidae* family (pigs) from the same taxonomic unit (*artiodactyla*) separated from the bovids and were closer to primates. Within the primates we further observed a slight but distinct separation of *cercopithecidae* (crab-eating and rhesus monkeys) from *hominidae* (human). Separation of the human breast milk samples is mostly explained by population factor (supplementary *fig. S1*). Commercial baby formula formed its own cluster, except the three formulas that appear to be close to bovids (*fig. 5*).

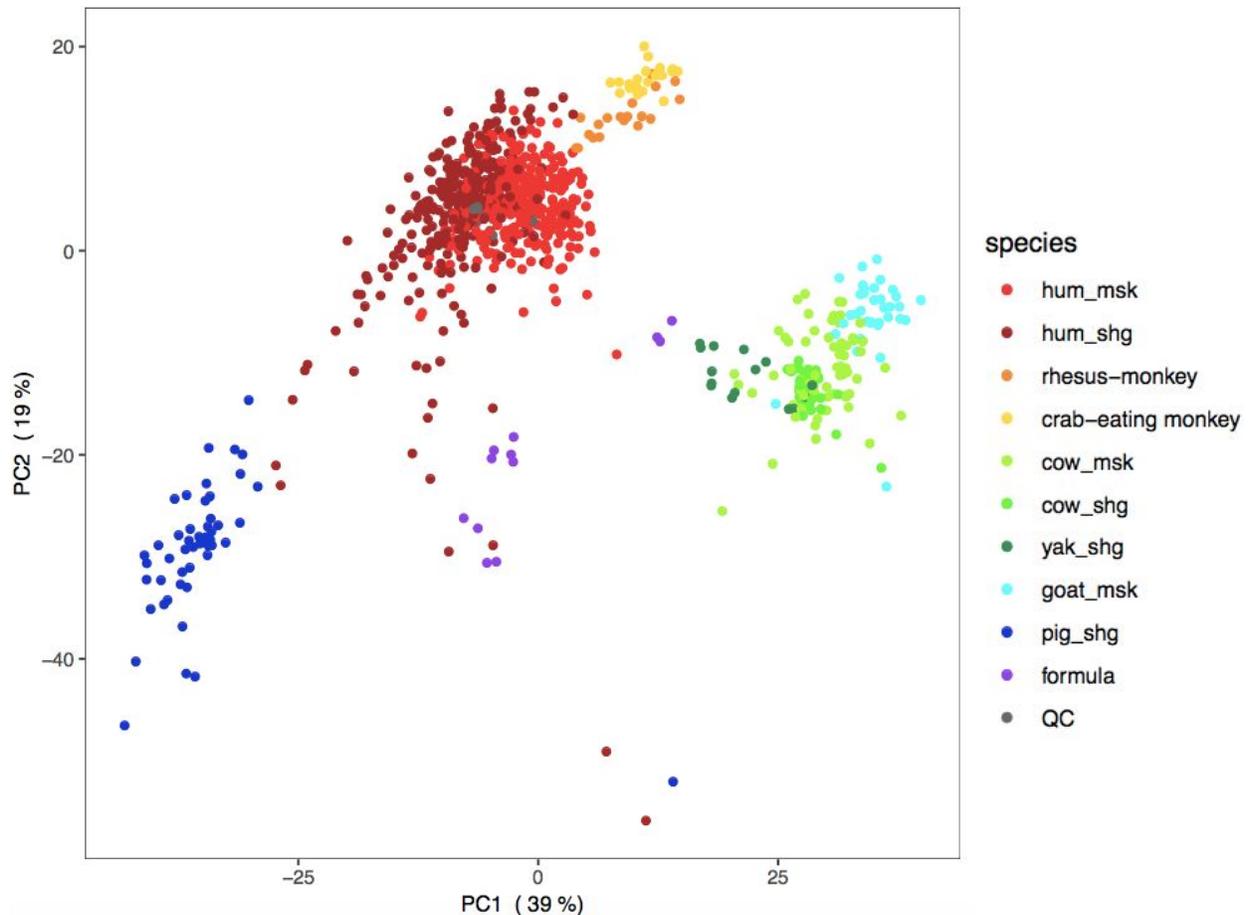


Fig. 5. Total milk lipidome variation. Principal component analysis for all samples based on concentrations of 647 detected lipid features. Colors represent species, human populations, and commercial baby formula.

Analysis of the hydrolyzed milk data revealed similar inter-species variation patterns: primates, bovidae, suidae, and commercial baby formula lipids formed distinct clusters in the principal component analysis. Interestingly, cercopithecidae did not separate from hominidae samples, so overall milk FA composition can be described as primate-specific rather than human specific (*fig. 6*).

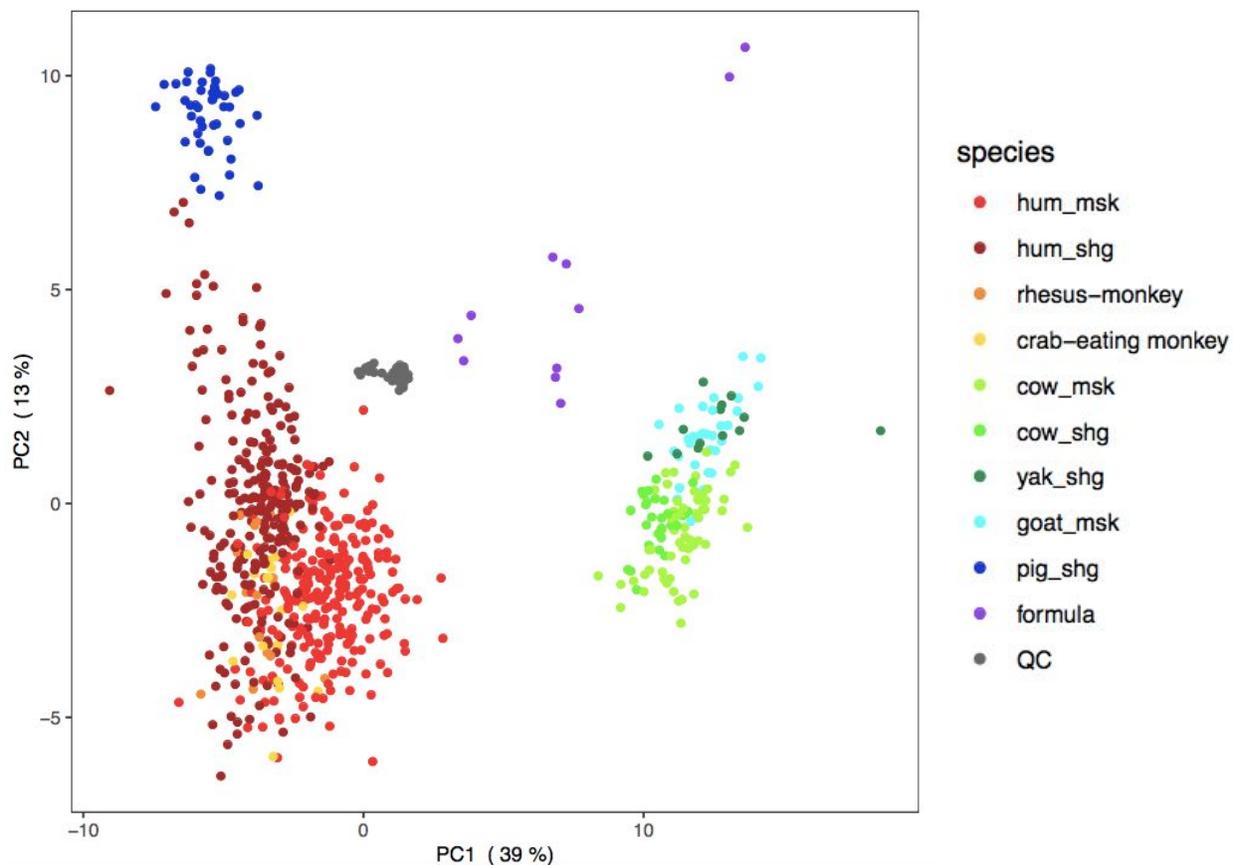
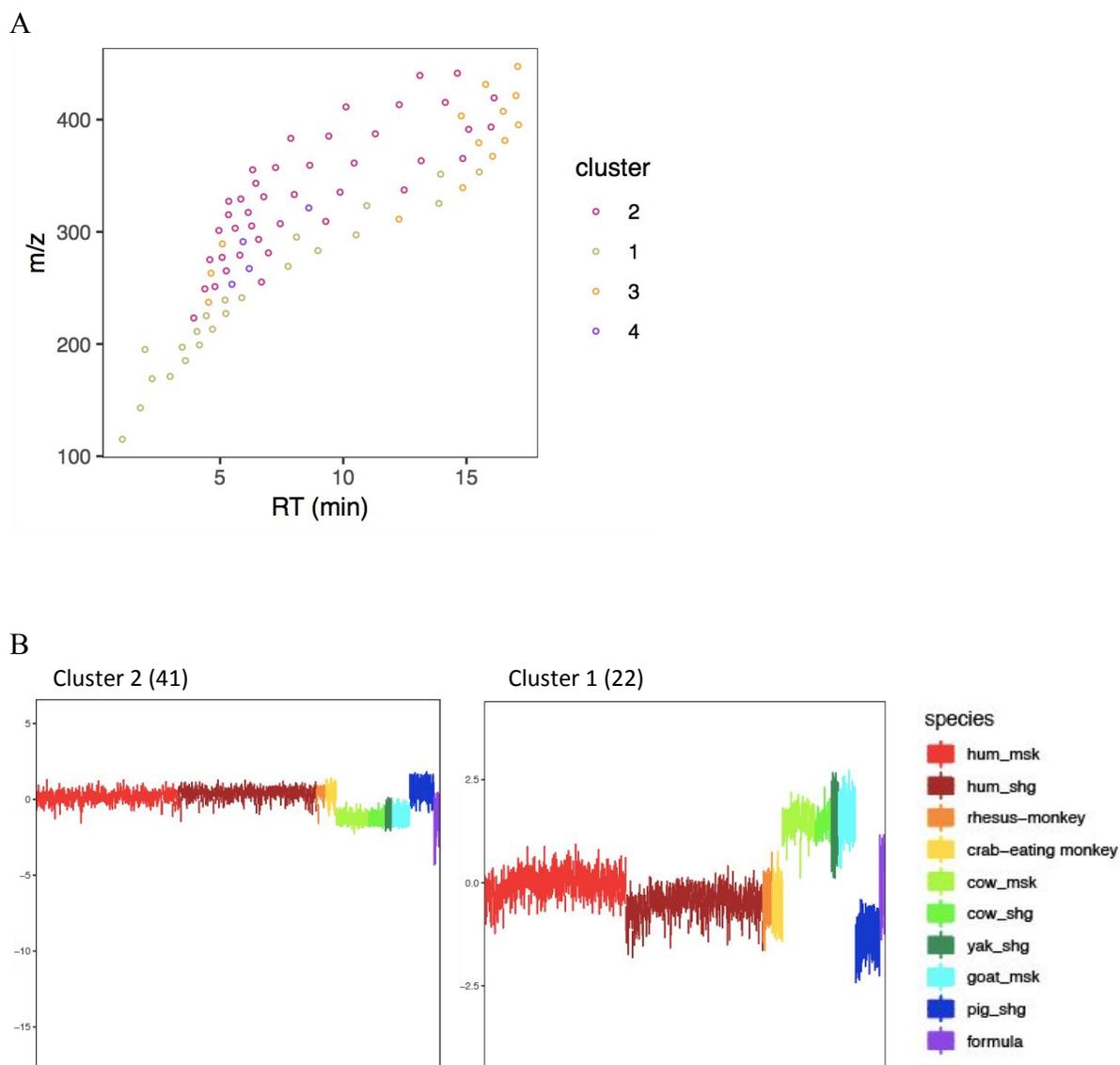


Fig. 6. Hydrolyzed milk lipidome variation. Principal component analysis for all samples based on concentrations of 81 annotated FAs. Colors represent species, human populations, and commercial baby formula.

Statistical analysis of the hydrolyzed milk lipidome identified 81 annotated FAs showing species-dependent differences in their feature intensity levels (ANOVA, BH-corrected $P < 0.05$). Unsupervised clustering of these 81 FAs based on their intensity profiles across samples yielded 4 clusters (**fig. 7A**). Long-chain FAs with different levels of unsaturation (cluster 2) show high levels of signal intensity in human, macaque, and pig samples. Milk formula FAs that belong to this cluster have signal intensity similar to that of bovidae. Short and medium-chain and predominantly saturated FAs (cluster 1) show high levels of signal intensity in bovidae - cow,

yak and goat samples and also in baby formula; human and macaque samples have low levels of short and medium-chain saturated FAs. Interestingly, in two of the four clusters the FA intensity levels in Russian population were lower than in Chinese population. Several long-chain FAs (cluster 4) including rare odd-chain ones, were higher in cow and yak samples. Long and predominantly odd chain FAs (cluster 3) showed higher signal intensities in goat and yak samples.



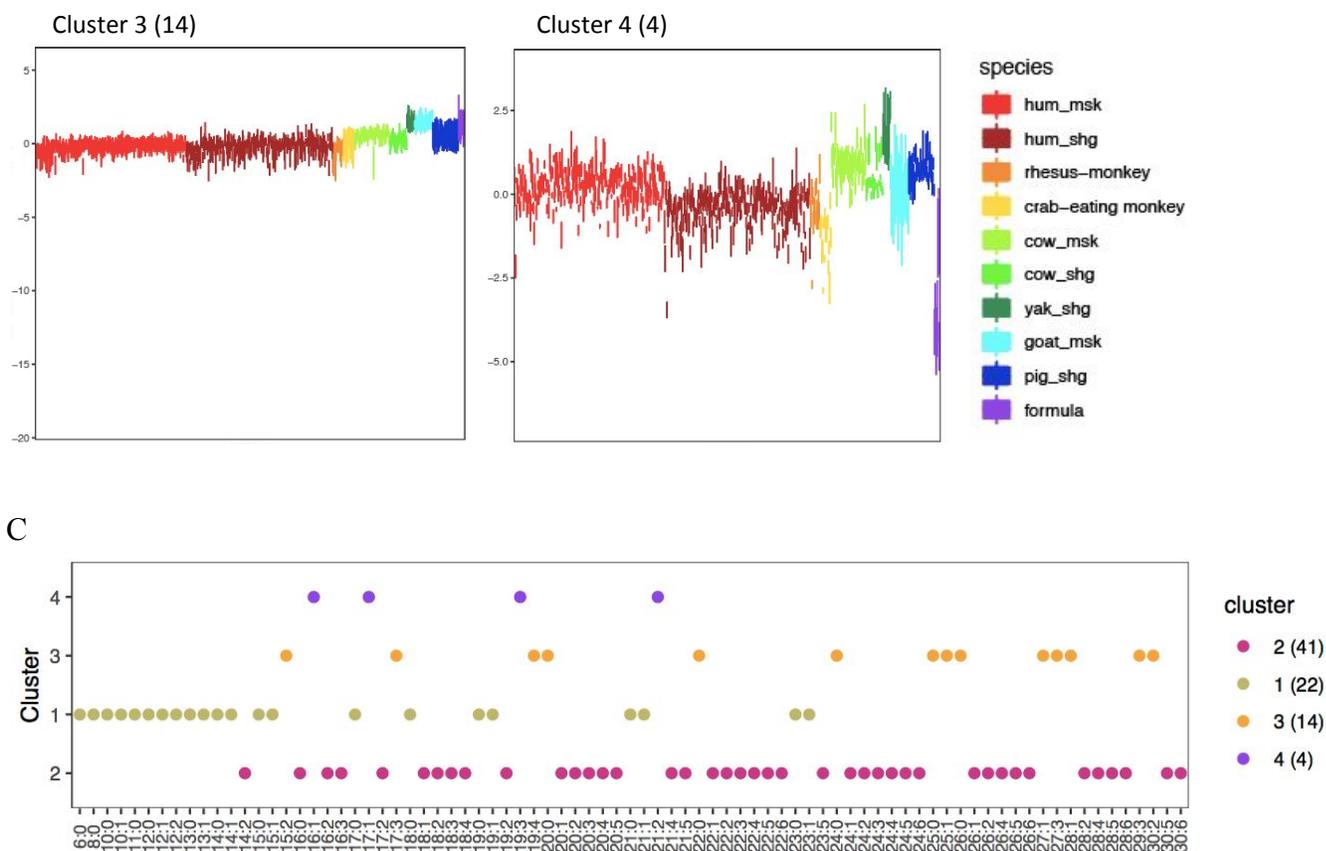
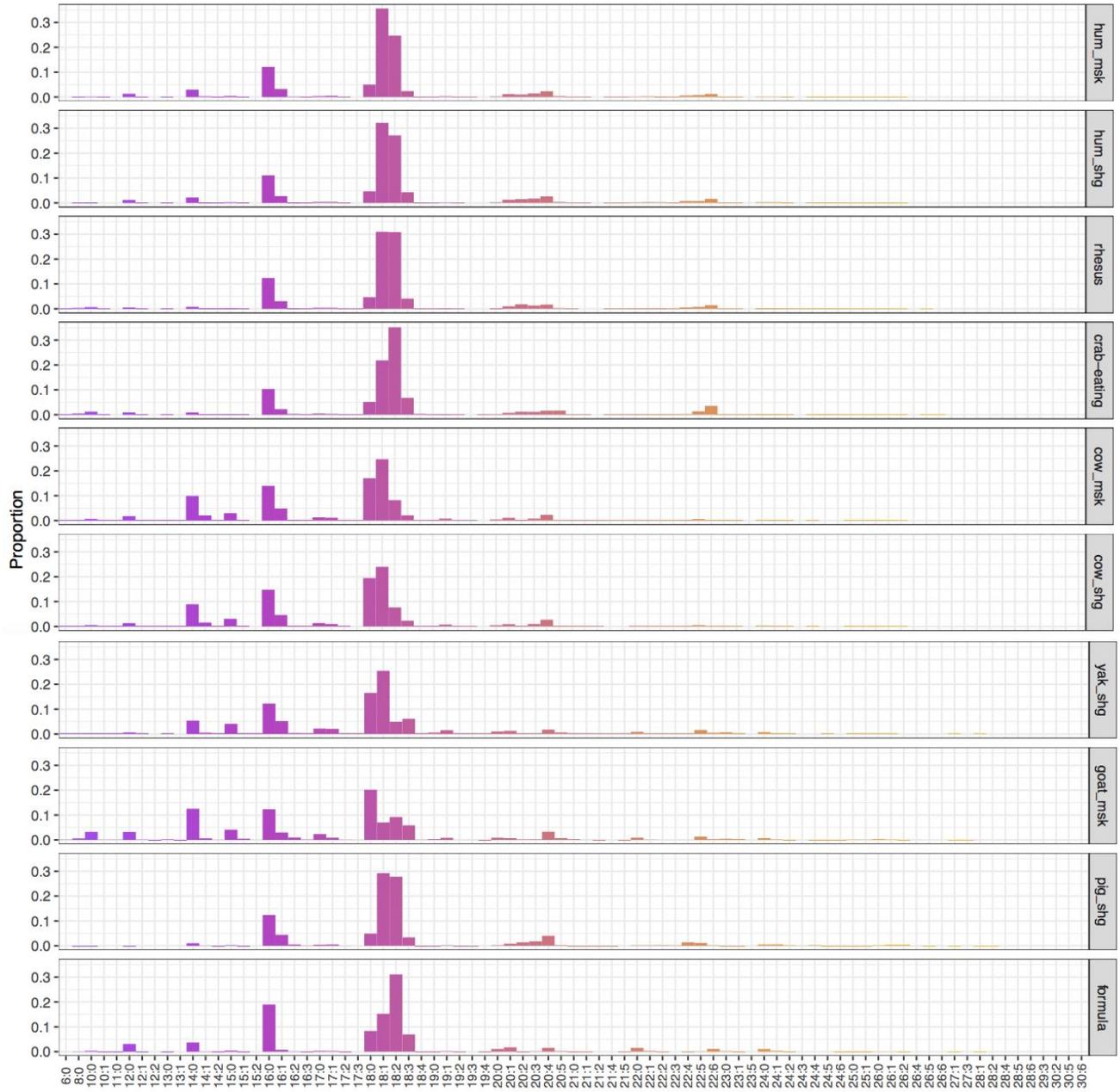


Fig. 7. Unsupervised cluster analysis of annotated FAs in milk. (A) 81 annotated FA peaks coloured by clusters. The x-axis represents the mass of the compound, m/z. The y-axis represents retention time of the compound in minutes, RT (min). (B) Distribution of signal intensities of FAs for a given cluster. Each box represents one sample; a number of features is indicated at the top, samples colored and ordered by species. (C) Distribution of FAs by clusters. The x-axis represents annotated FAs ordered by the chain length and then by the number of double bonds. The y-axis represents clusters. Colored by clusters.

In order to assess the content of each FA, we calculated proportions as: mean intensity of particular FA divided by the sum of mean intensities of all other annotated FAs for each species separately (**fig 8A**). The most abundant FAs in milk are 14:0, 16:0, 18:0, 18:1, 18:2, 18:3, with low levels of 20:4 and 22:6, and trace amounts of the rest. In addition to considering each species

separately, we joined our samples into groups, representing primates (human and macaque samples), bovines (cow and yak samples), goat samples separately, and pid samples (*fig. 8B*).

A



B

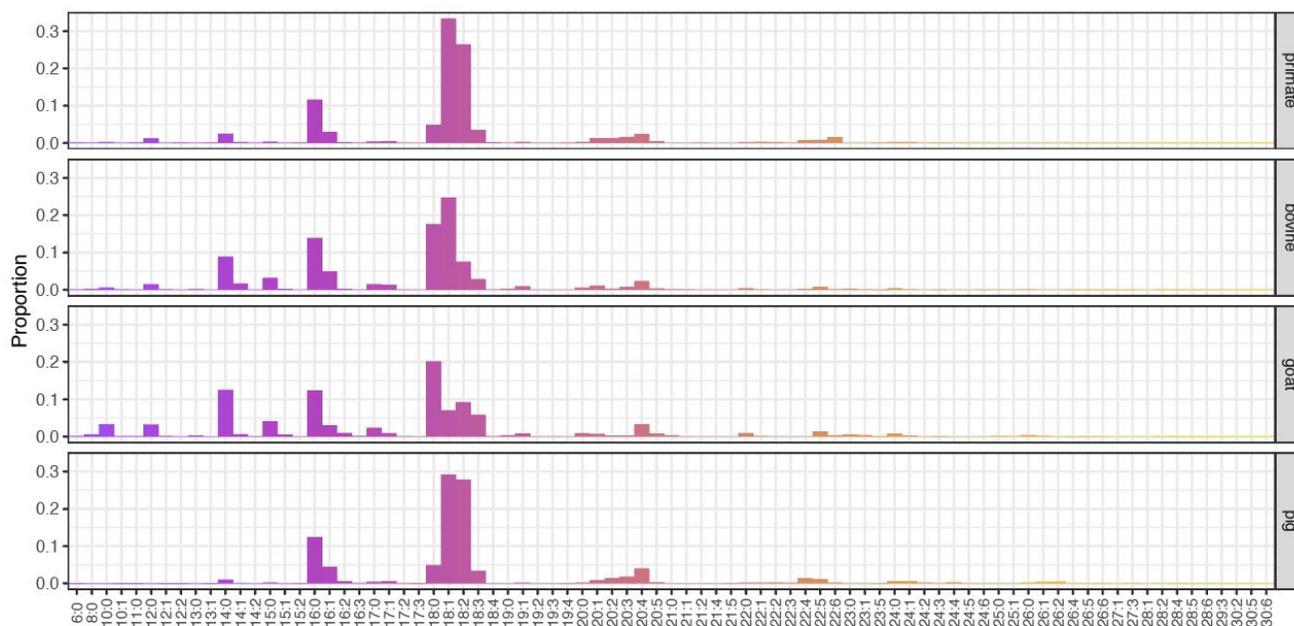
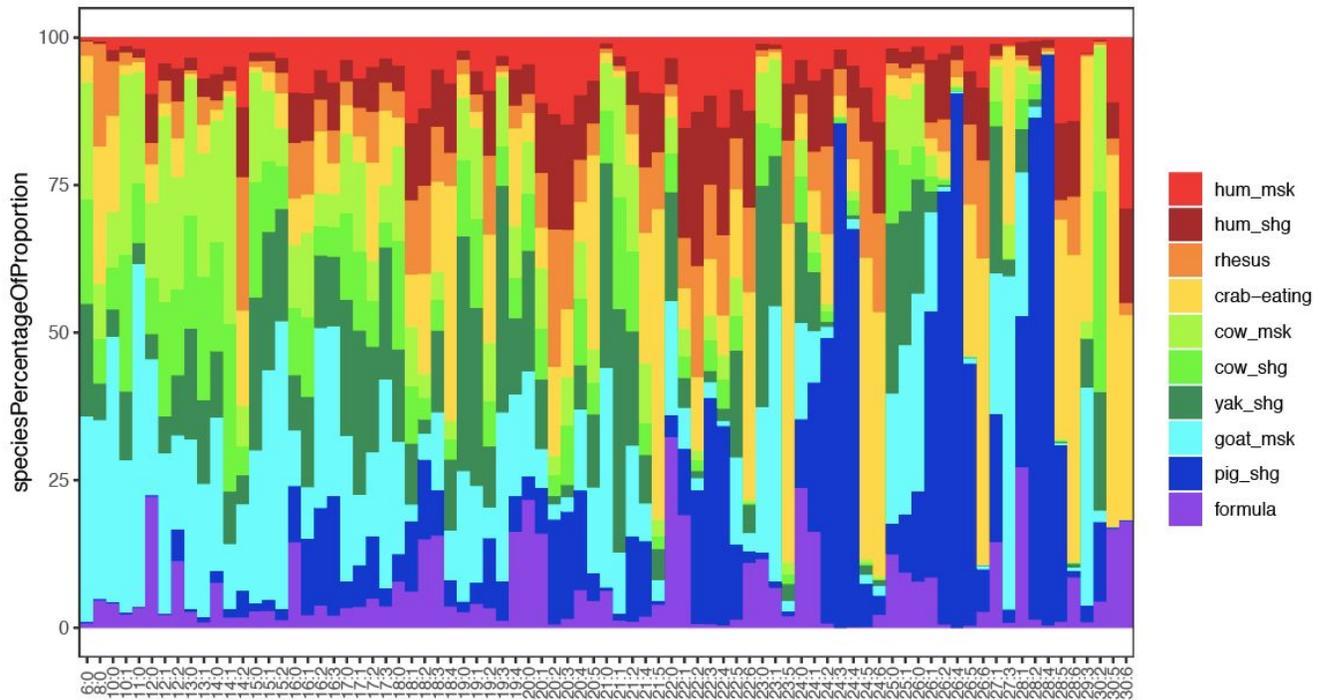


Fig. 8. Proportions of milk FAs in each species and in species' groups. (A) Distribution of the calculated proportions of the signal intensities of the annotated FAs separately for each species. The x-axis represents annotated FAs ordered by the chain length and then by the number of double bonds. The y-axis represents proportion from the total pool of annotated FAs. (B) Distribution of the calculated proportions of the signal intensities of the annotated FAs for the species, joined into 4 groups. The x-axis represents annotated FAs ordered by the chain length and then by the number of double bonds. The y-axis represents proportion from the total pool of annotated FAs.

In agreement with results obtained in a test sample set, goat milk samples were high in saturated FAs: 14:0, 16:0 and 18:0, but low in unsaturated ones. By contrast, pig milk samples were high in mono- and polyunsaturated species: 18:1 and 18:2. The milk of bovinds was in between goat and pig milk - it contained high proportions of saturated 18:0, for unsaturated species 18:1 and 18:2 they are higher than in goats but lower than in pigs. Primate milk samples showed low levels of saturated, and high levels of unsaturated FAs.

In order to assess the presence of very long-chain FAs that are found in low amounts in milk, we calculated the percentage of FAs' proportions in the milk content for each particular species as: the proportion of a particular FA divided by the sum of the calculated proportions of this FA in all species and multiplied by 100% (*fig. 9A*).

A



B

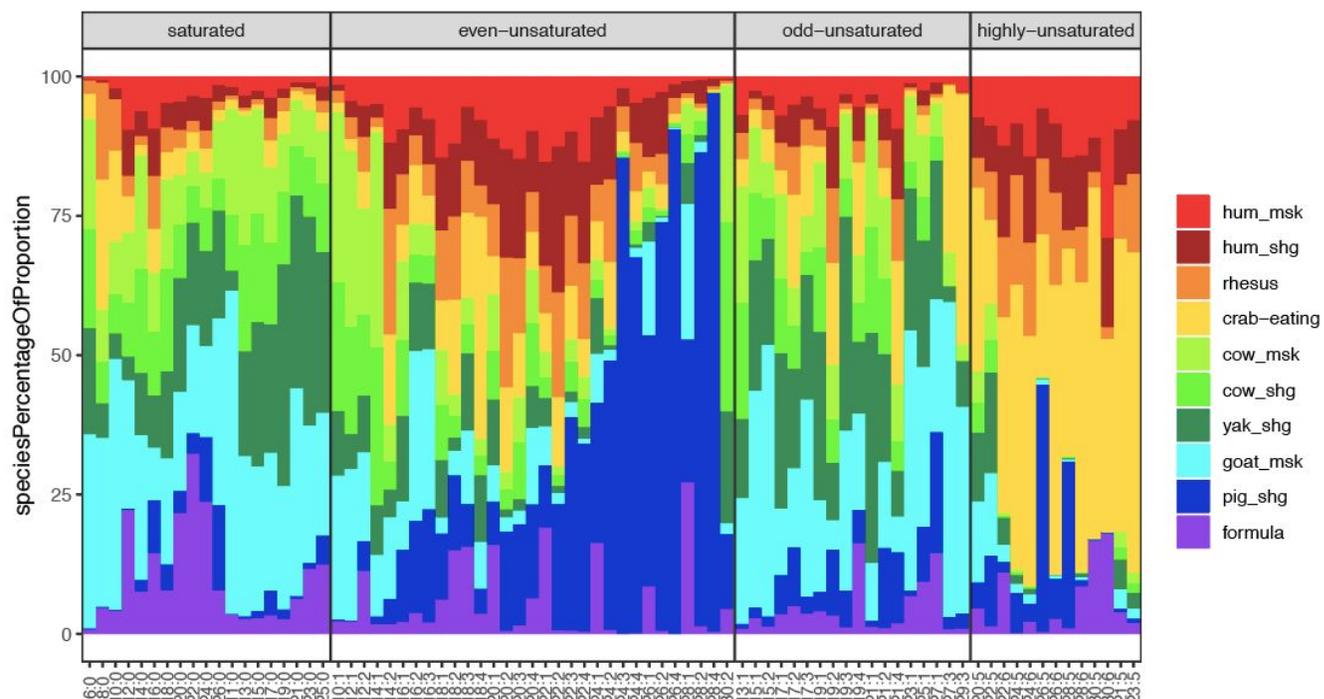
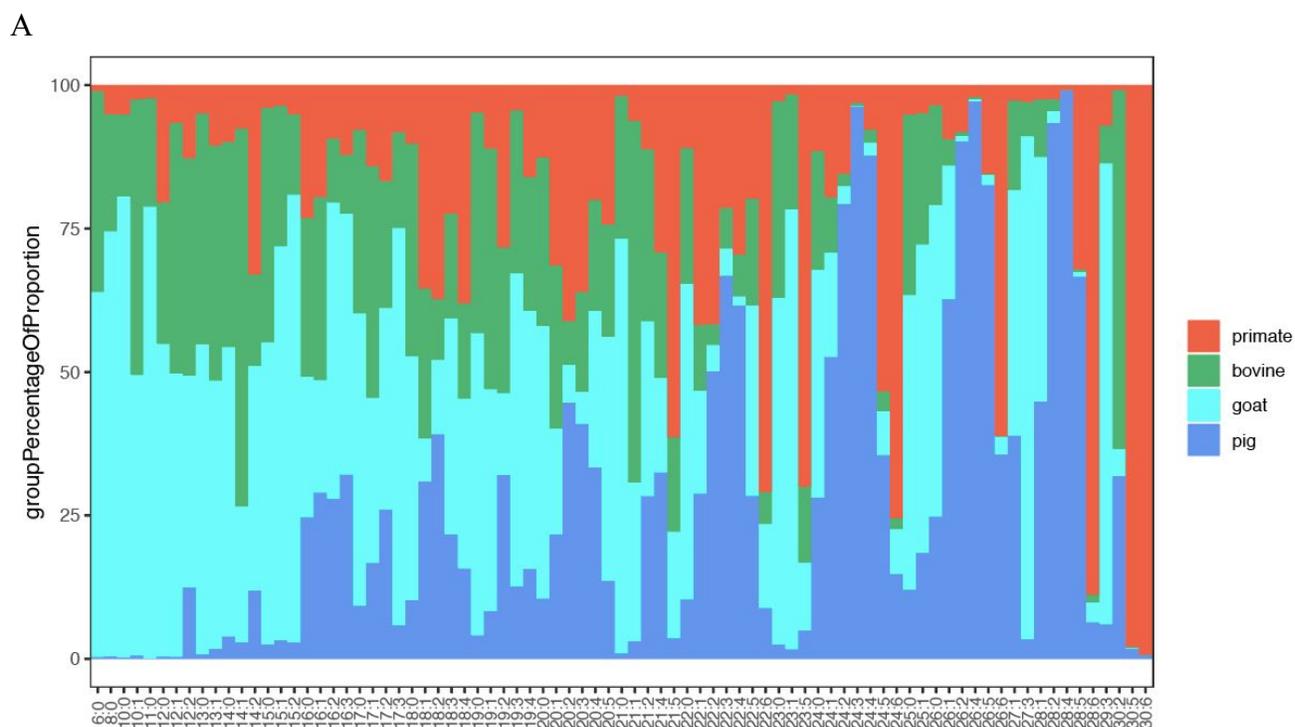


Fig. 9. Percentage of the milk FAs proportions across species. (A) Distribution of the calculated percentage of the proportion of each FA across all species. The x-axis represents annotated FAs ordered by the chain length and then by the number of double bonds. The y-axis represents the percentage of the FA proportion in each species. (B) Distribution of the calculated percentage of the proportion of each FA across all species. The x-axis represents annotated FAs grouped by the number of double bonds and by the parity of the chain length. The y-axis represents the percentage of the FA proportion in each species.

Grouping the FAs according to carbon chain and number of double bonds resulted in four groups: saturated FAs (even- and odd-chain, with 0 double bonds), unsaturated even-chain (containing 1,2,3 or 4 double bonds), unsaturated odd-chain (containing 1,2,3 or 4 double bonds), and highly unsaturated (containing 5 or 6 double bonds). Based on this FA separation, the species-specific differences became more clear. Saturated FAs and odd-chain unsaturated ones were high in bovidae species (cow, yak and goat). Unsaturated medium even-chain FAs

were similarly high in bovidae species. Unsaturated long even-chain FAs were high in primate samples. Lastly, unsaturated very long even-chain FAs were high in pigs. Notably, if we consider long-chain FAs with higher number of double bonds (5 or 6), we see that those are mostly present in primate samples (*fig. 9B*).

In order to evaluate FA abundances specific for groups or families of species, we joined all samples into four groups representing primates (human and macaque samples), bovines (cow and yak samples), goat samples separately, and pig samples (*fig. 10A*), and also group FAs according to carbon chain and number of double bonds into four groups, as described above (*fig. 10B*).



B

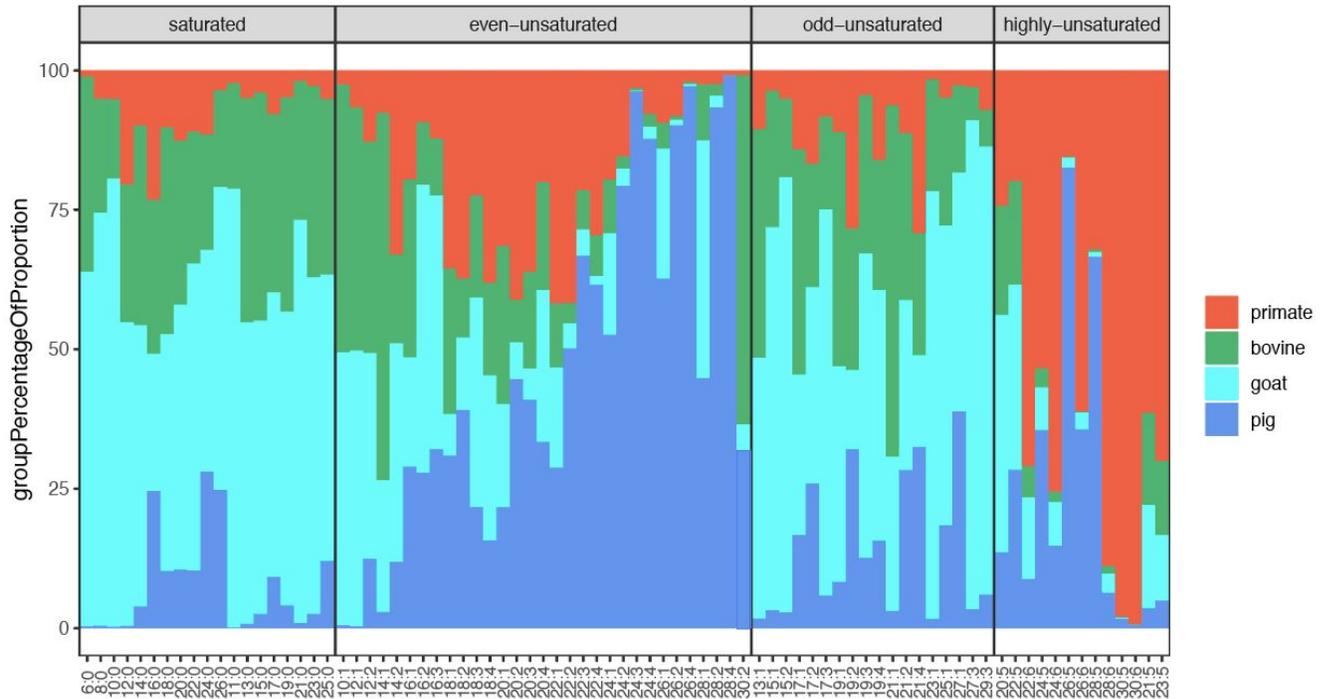
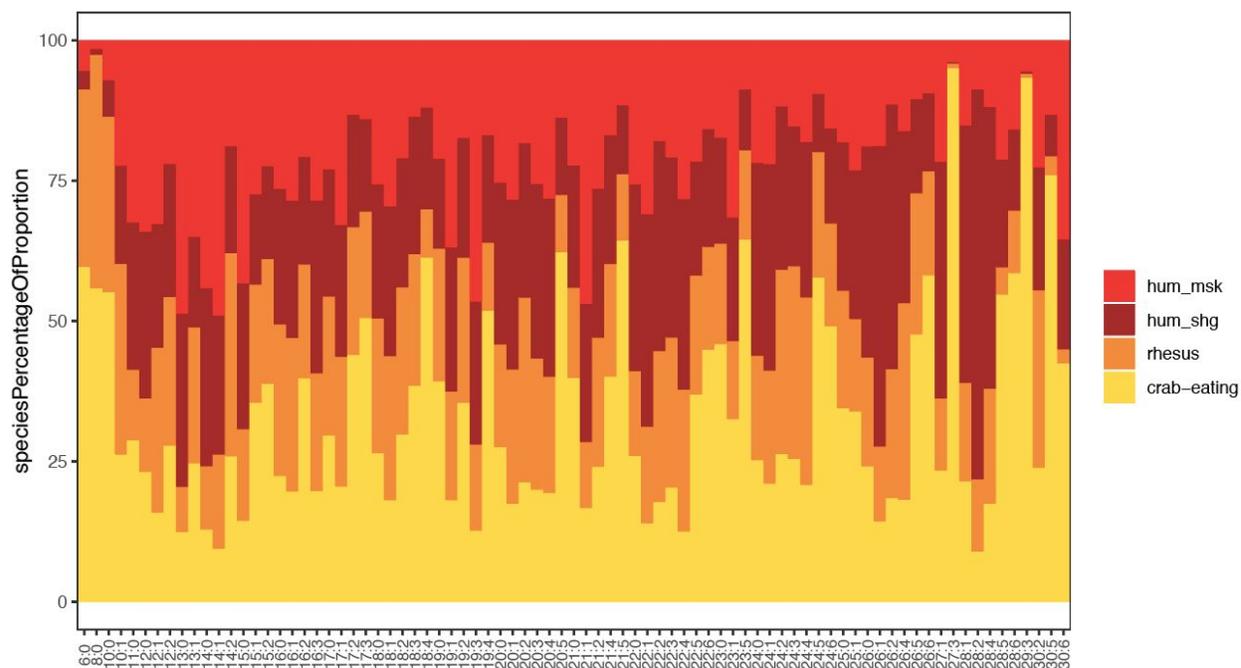


Fig. 10. Percentage of the FAs proportions across groups of species in milk. (A) Distribution of the calculated percentage of the proportion of each FA across four groups of species. The x-axis represents annotated FAs ordered by the chain length and then by the number of double bonds. The y-axis represents the percentage of the FA proportion in each group of species. (B) Distribution of the calculated percentage of the proportion of each FA across four groups of species. The x-axis represents annotated FAs grouped by the number of double bonds and by the parity of the chain length. The y-axis represents the percentage of the FA .

If we consider primate samples separately, and calculate the percentage of the proportions of FAs for each primate species as: the proportion of a particular FA divided by the sum of the calculated proportions of this FA in primate species and multiplied by 100%, we can evaluate the relative abundance of FAs across investigated primate species and two human populations. Clustering the FAs into four groups demonstrates that even-chain unsaturated FAs were higher in humans, and in particular, in Chinese population (*fig. 11*).

A



B

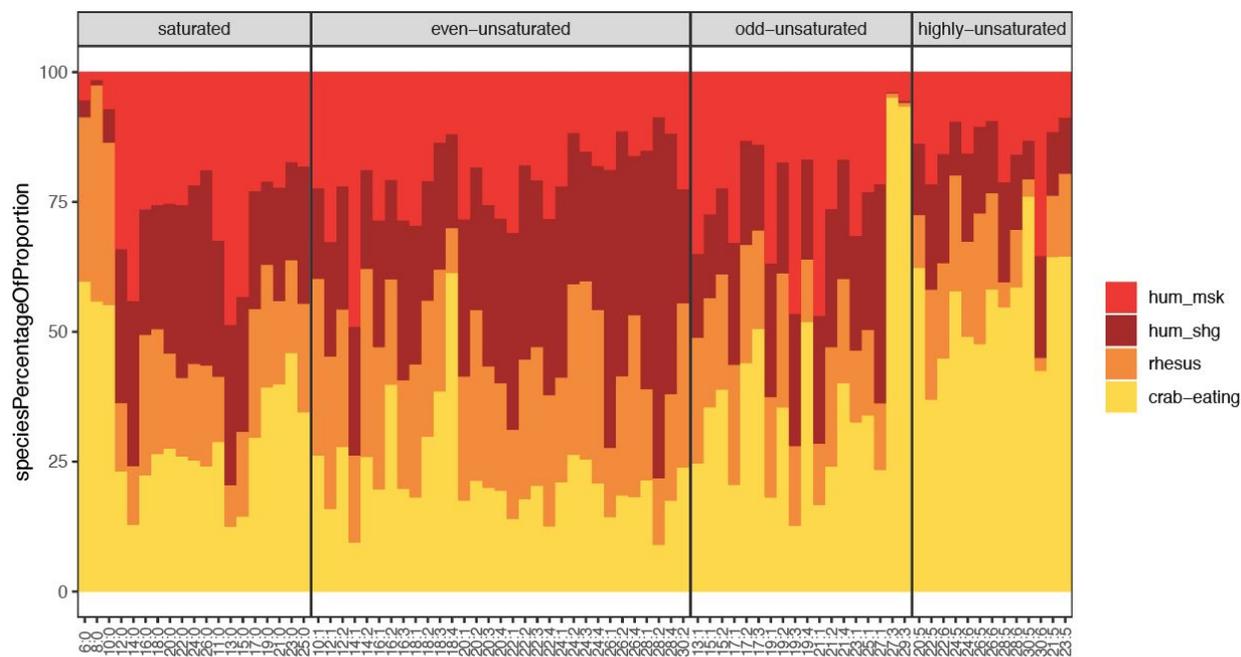


Fig. 11. Percentage of the FAs proportions across primates in milk. (A) Distribution of the calculated percentage of the proportion of each FA across primates. The x-axis represents annotated FAs ordered by the chain length and then by the number of double bonds. The y-axis

represents the percentage of the FA proportion in each primate species. (B) Distribution of the calculated percentage of the proportion of each FA across primate species. The x-axis represents annotated FAs grouped by the number of double bonds and by the parity of the chain length. The y-axis represents the percentage of the FA.

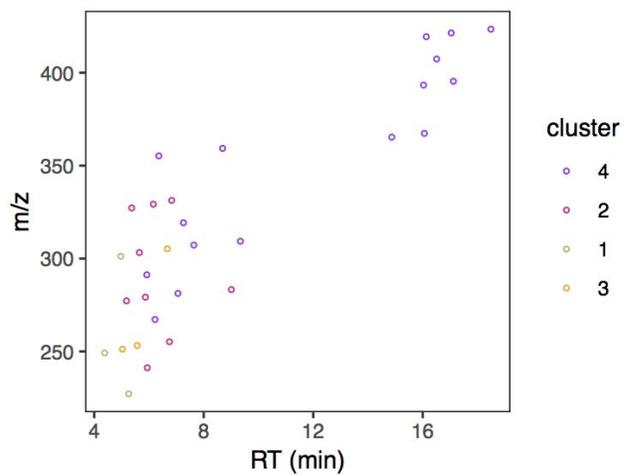
3.3 Analysis of the brain lipidome FA composition

To assess the relationship between differences in FA levels among species and the differences in the FA composition of the developing baby brain, we analyzed FA composition of the brain lipidome in 4 species contained within the milk dataset: human, macaque, goat and pig, and also in chimpanzees. For each species, we focused on the gray matter dissected from the prefrontal cortical region (PFC) and cerebellar gray matter (CB) of the brain. All samples were taken from well preserved postmortem tissue with short postmortem delay. The individuals' age was limited to the first year of postnatal lifespan (supplementary *tbl. 3*). In total, we assessed the brain lipidome FA composition in 194 samples, of them humans - 92, chimpanzees - 18, macaques - 38, pigs - 20, and goats - 26.

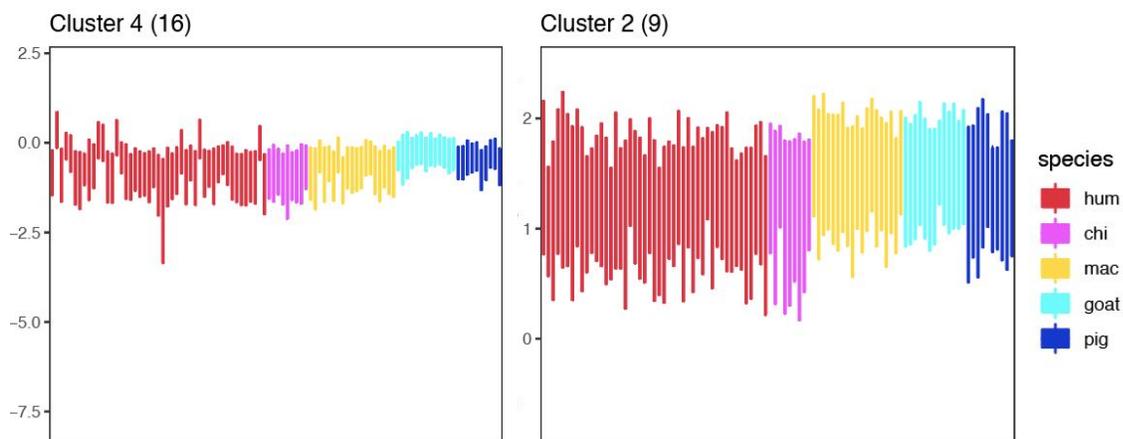
The lipid extraction, hydrolysis, FA intensity measurements, intensity normalization, and FA annotation were carried out using the same procedures, as for the milk lipidome. The analysis of the annotated FAs from the brain revealed 31 FAs with species-specific differences in prefrontal cortex (ANOVA, BH-corrected $P < 0.05$). Unsupervised clustering of these 31 FAs based on their intensity profiles across samples yielded four clusters (*fig. 12A*). Long FAs, including odd-chain ones, with different levels of saturation (cluster 4) were highly abundant in goat brain samples. Long-chain FAs including the most abundant in brain: stearic (16:0), oleic (18:0), linoleic (18:2), alpha-linolenic (18:3), arachidonic (20:4), docosatetraenoic (22:4),

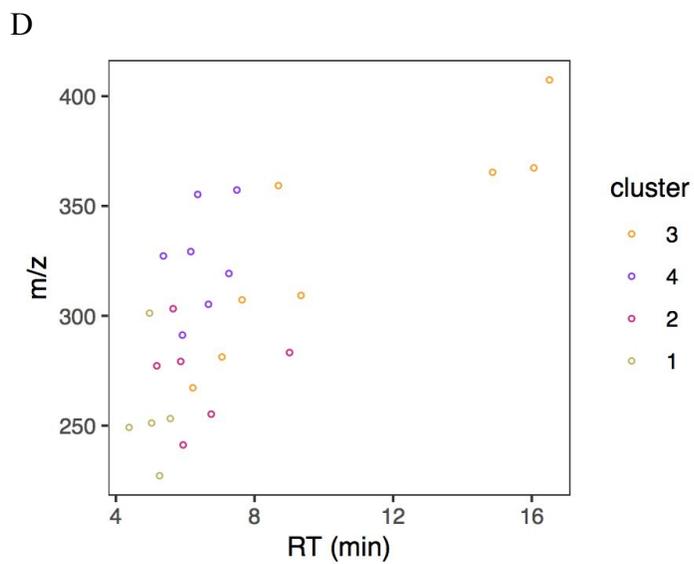
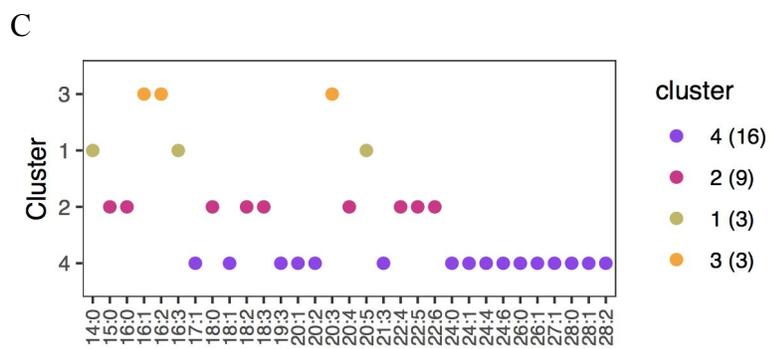
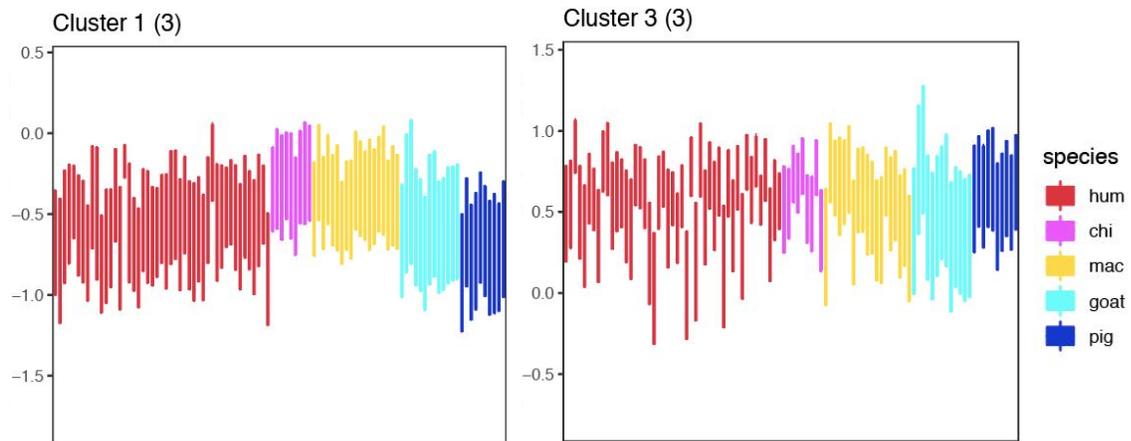
docosapentaenoic (22:5) and docosahexaenoic (22:6) FAs (cluster 2) showed high levels of signal intensity in macaque, goat and pig samples.

A



B





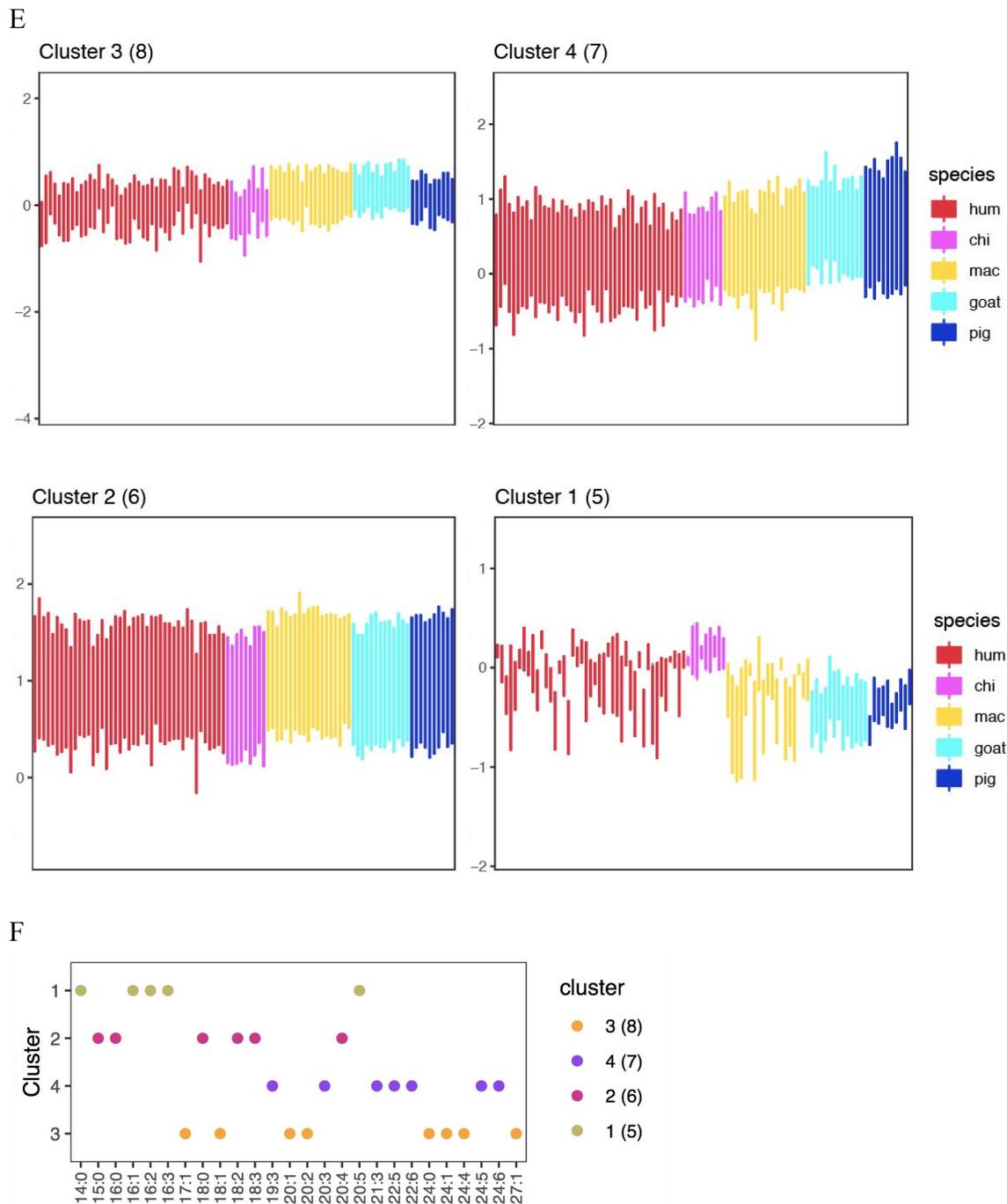


Fig. 12. Cluster analysis of annotated FAs in the brain. *Prefrontal cortex:* (A) 31 annotated FA peaks in PFC coloured by clusters. The x-axis represents the mass of the compound, m/z. The y-axis represents retention time of the compound in minutes, RT (min). (B) Distribution of signal intensities of FAs for a given cluster in PFC; each box represents one sample; number of features is indicated at the top, samples coloured and ordered by species. (C) Distribution of FAs by

clusters in PFC. The x-axis represents annotated FAs ordered by the chain length and then by the number of double bonds. The y-axis represents clusters. Colors indicate clusters. *Cerebellum:* (D) 31 annotated FA peaks in CB coloured by clusters. The x-axis represents the mass of the compound, m/z. The y-axis represents retention time of the compound in minutes, RT (min). (E) Distribution of signal intensities of FAs for a given cluster in CB; each box represents one sample; number of features is indicated at the top, samples coloured and ordered by species. (F) Distribution of FAs by clusters in CB. The x-axis represents annotated FAs ordered by the chain length and then by the number of double bonds. The y-axis represents clusters. Colors indicate clusters.

Proportions of FAs were calculated as: the mean intensity of particular FA divided by the sum of mean intensities of all other annotated FAs for each species separately (*fig. 13*). Brain FA composition as opposed to milk, was present by fewer FAs. In addition to highly abundant in milk 14:0, 16:0, 18:0, 18:1, 18:2 and 18:3, we detected high levels of omega-6 arachidonic (20:4) FA, and omega-3 docosahexaenoic FA (22:6), as well as 22:6 precursor - docosapentaenoic acid (22:5), and the product of 20:4 elongation - docosatetraenoic acid (22:4), which goes in agreement with the previous knowledge. Interestingly, levels of AA (20:4) are high in all species, while DHA (22:6) was higher in goat samples in the prefrontal cortex; for cerebellum, levels of DHA were high in goat and macaque samples.

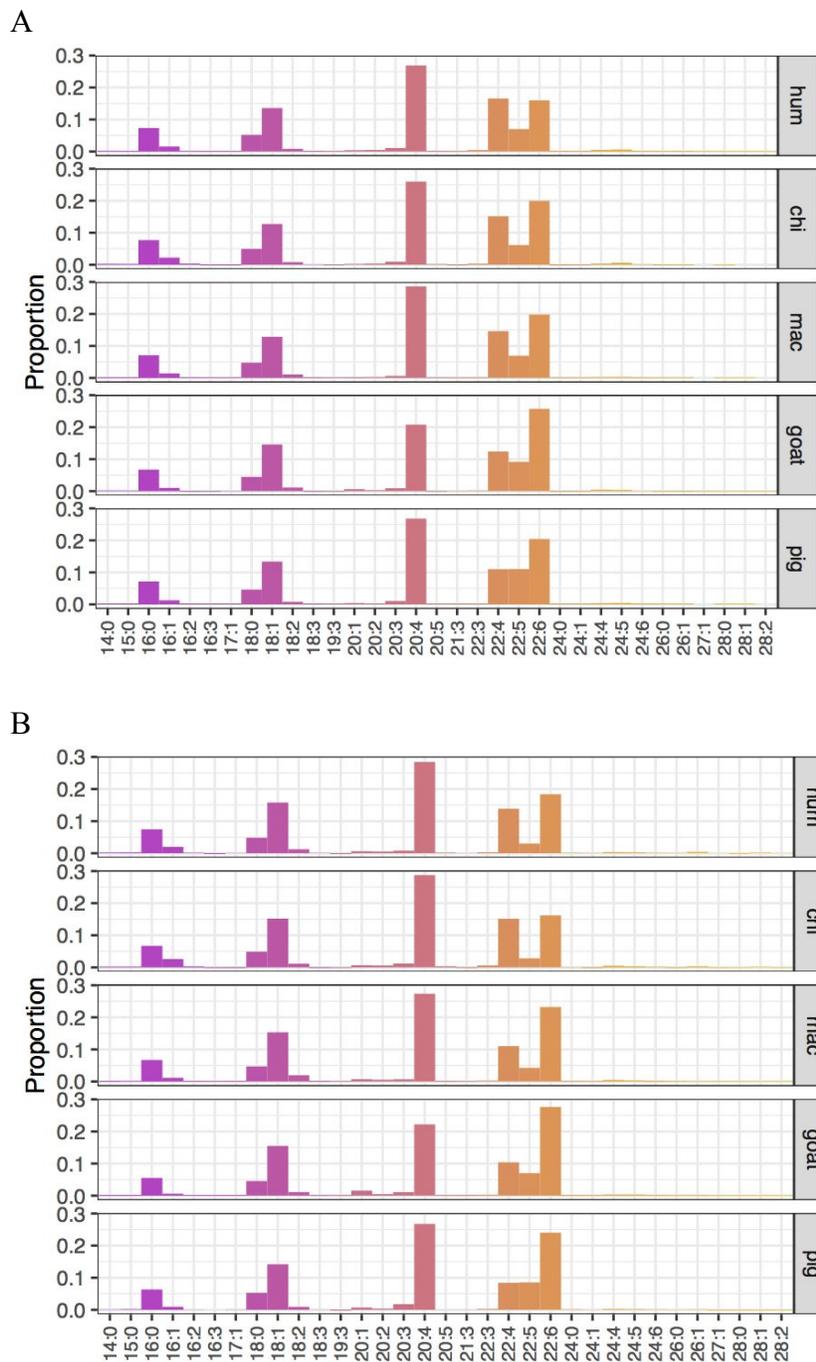


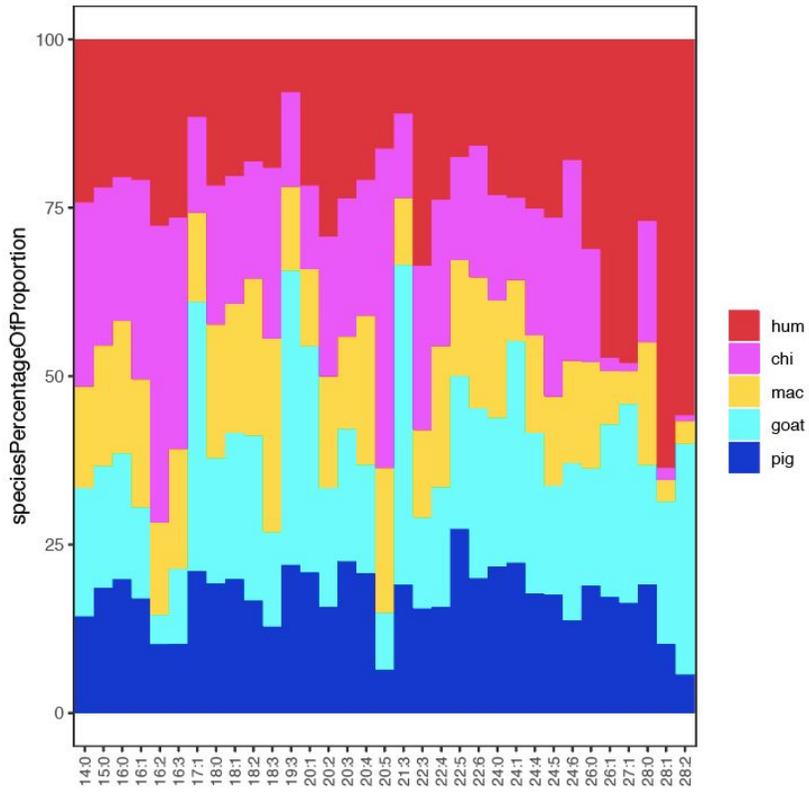
Fig. 13. Proportions of FAs in each species in the brain. (A) Distribution of the calculated proportions of the signal intensities of the annotated FAs separately for each species in PFC. The x-axis represents annotated FAs ordered by the chain length and then by the number of double bonds. The y-axis represents proportion from the total pool of annotated FAs. (B) Distribution of the calculated proportions of the signal intensities of the annotated FAs separately for each

species in CB. The x-axis represents annotated FAs ordered by the chain length and then by the number of double bonds. The y-axis represents proportion from the total pool of annotated FAs.

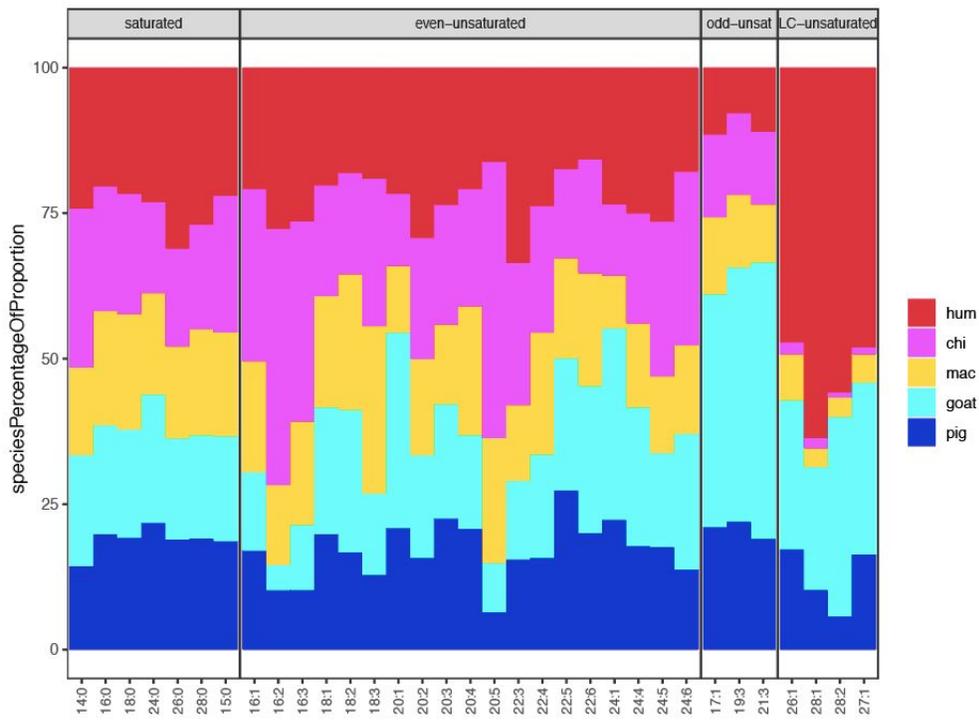
Percentage of the proportions of FAs in the brain for each species was calculated as: proportion of the particular FA divided by the sum of the proportions of this FA in all species and multiplied by 100%. This representation gives us a better understanding of what species possess higher intensities of particular FAs in the brain, including very low abundant ones (*fig.14A* for PFC; *fig. 14C* for CB).

We then combined all FAs into four groups: saturated with 0 double bonds; even-chain unsaturated FAs containing 1, 2, 3, 4, 5 or 6 double bonds; odd-chain unsaturated FAs with 1 or 3 double bonds (there were only three odd-chain FAs in our dataset - 17:1, 19:3 and 21:3); and long-chain unsaturated FAs, with chain length from 26 to 28, the highest that we could annotate in brain data, with 1 or 2 double bonds. Interestingly, the proportion of the FAs from this last group was higher in humans than in other species (*fig.14B* for PFC; *fig. 14D* for CB).

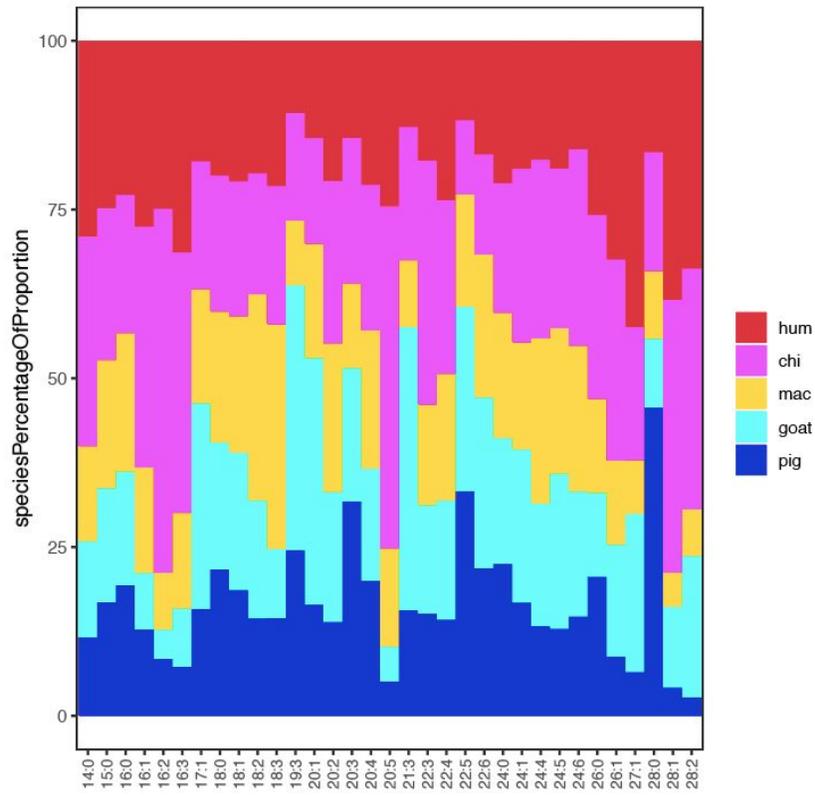
A



B



C



D

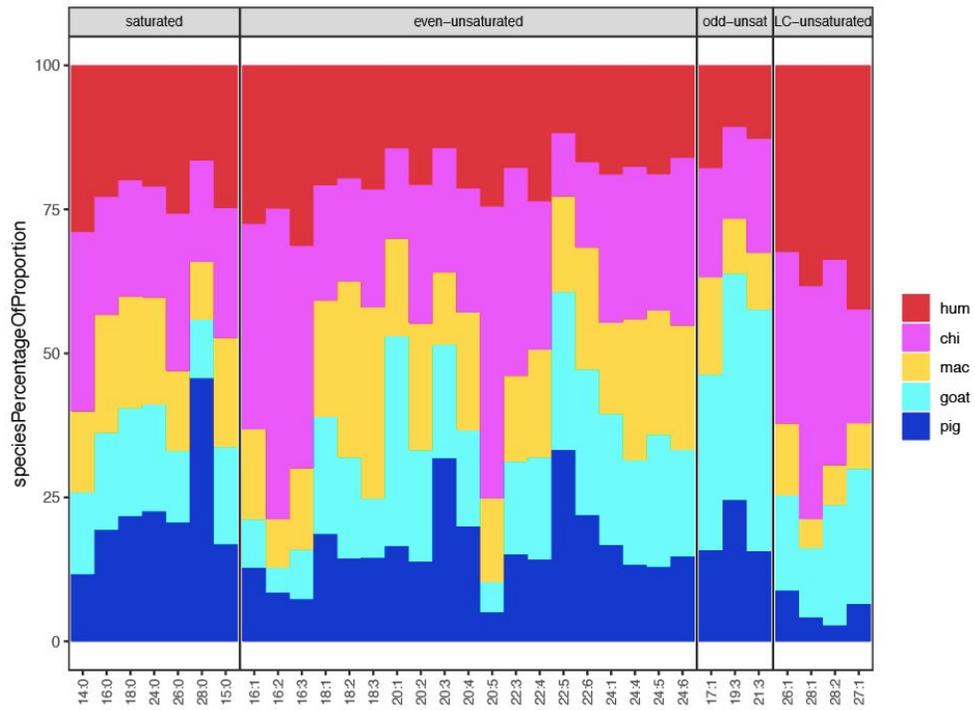
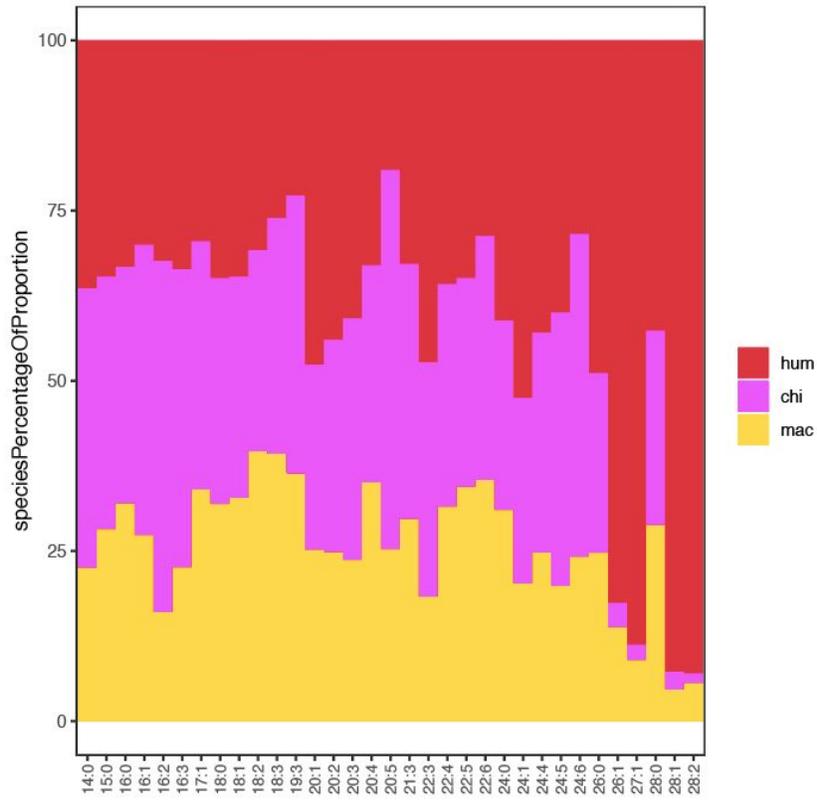


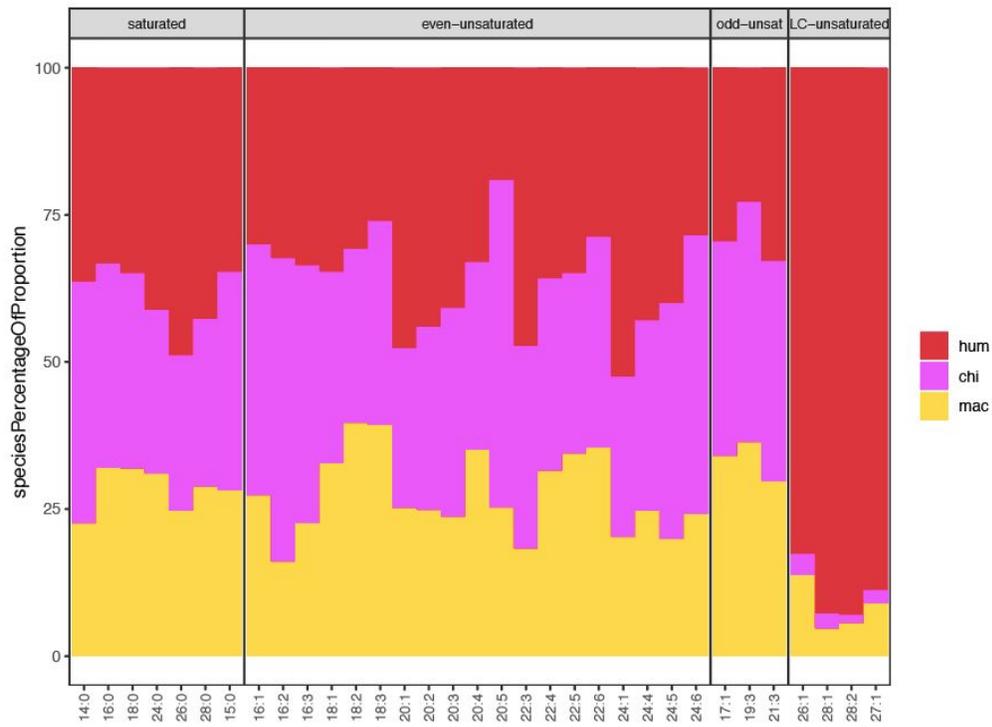
Fig. 14. Percentage of the FAs proportions across species in the brain. *Prefrontal cortex:* (A) Distribution of the calculated percentage of the proportion of each FA across all species in PFC. The x-axis represents annotated FAs ordered by the chain length and then by the number of double bonds. The y-axis represents the percentage of the FA proportion in each species. (B) Distribution of the calculated percentage of the proportion of each FA across all species in PFC. The x-axis represents annotated FAs grouped by the number of double bonds and by the parity of the chain length. The y-axis represents the percentage of the FA proportion in each species. *Cerebellum:* (C) Distribution of the calculated percentage of the proportion of each FA across all species in CB. The x-axis represents annotated FAs ordered by the chain length and then by the number of double bonds. The y-axis represents the percentage of the FA proportion in each species. (D) Distribution of the calculated percentage of the proportion of each FA across all species in CB. The x-axis represents annotated FAs grouped by the number of double bonds and by the parity of the chain length. The y-axis represents the percentage of the FA proportion in each species.

If we consider primate samples separately and calculate the percentage of the proportions of FAs for each primate species as: calculated proportion of a particular FA divided by the sum of calculated proportions of this FA in all primate species and multiplied by 100%, and then join FAs into four groups as described above, we can see that FAs from the last group have the highest proportion in human samples (*fig. 15A* and *fig. 15B*). The effect was more pronounced in the prefrontal cortex than in cerebellum (*fig. 15C* and *fig. 15D*).

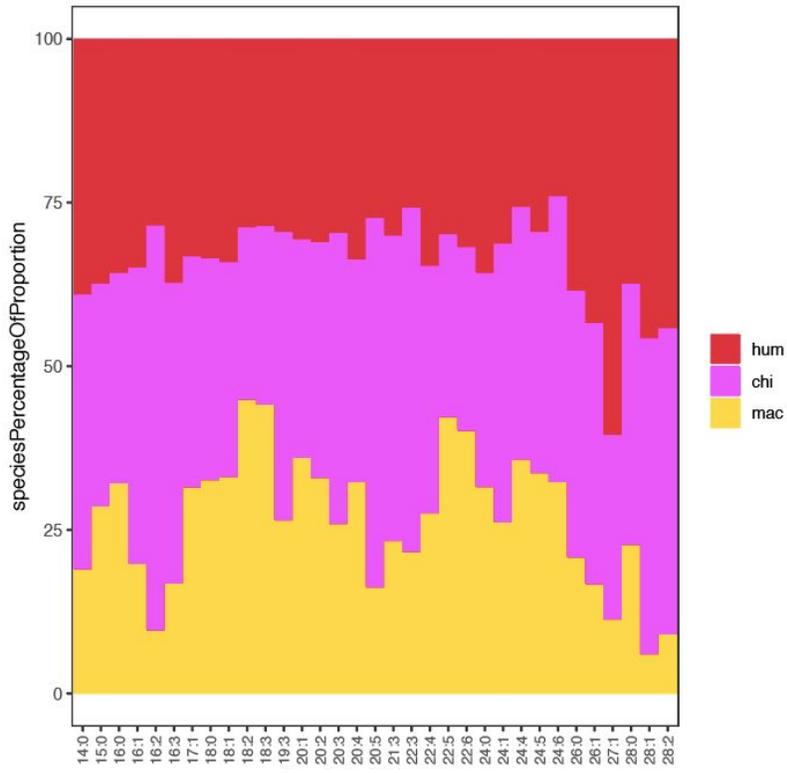
A



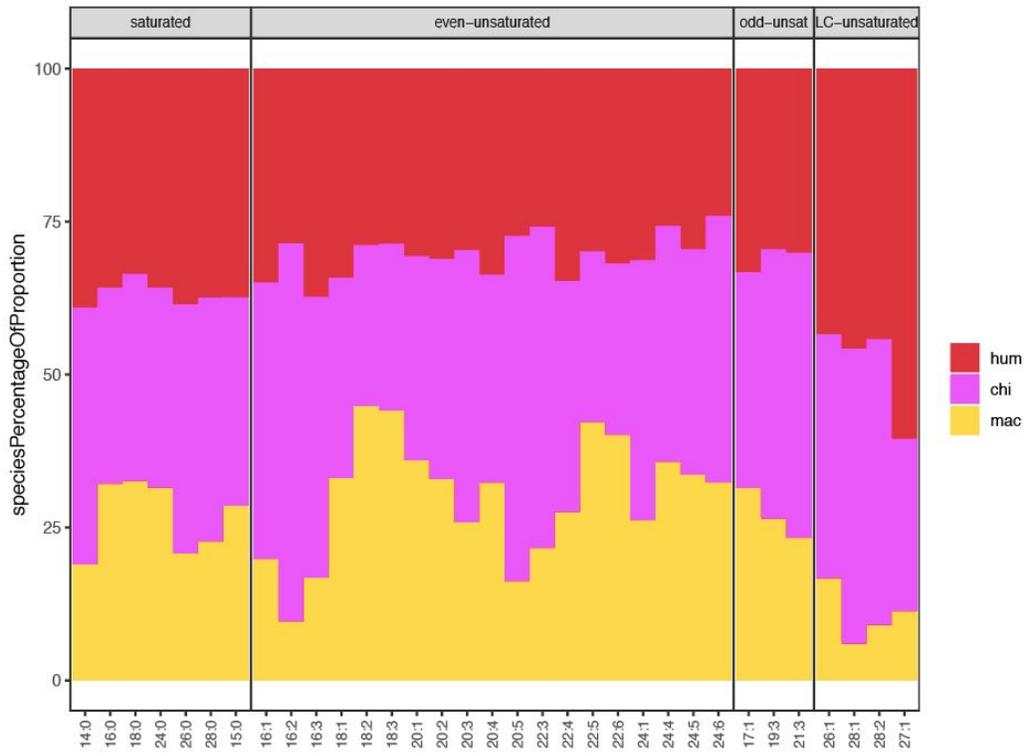
B

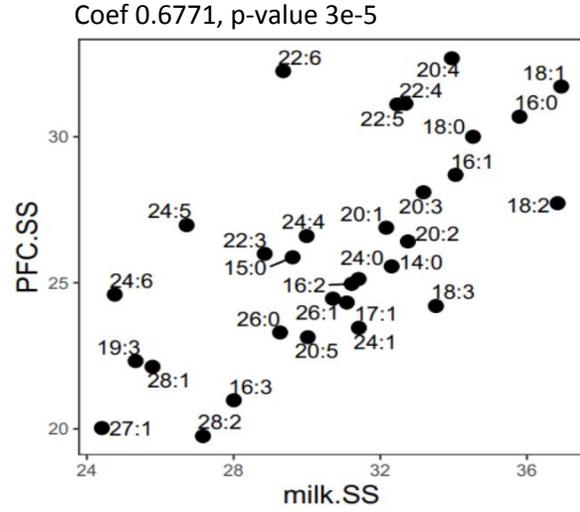
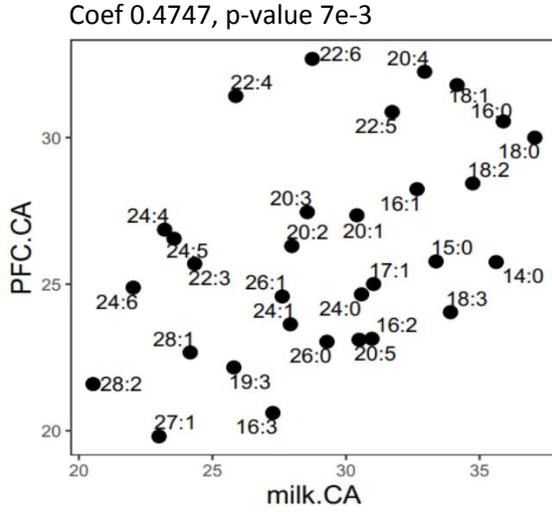


C



D





B

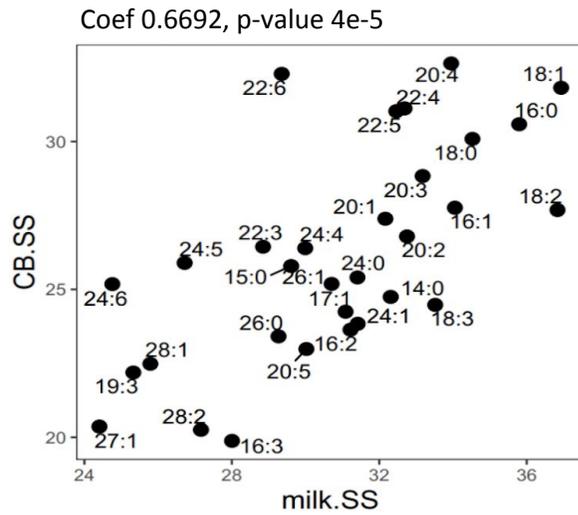
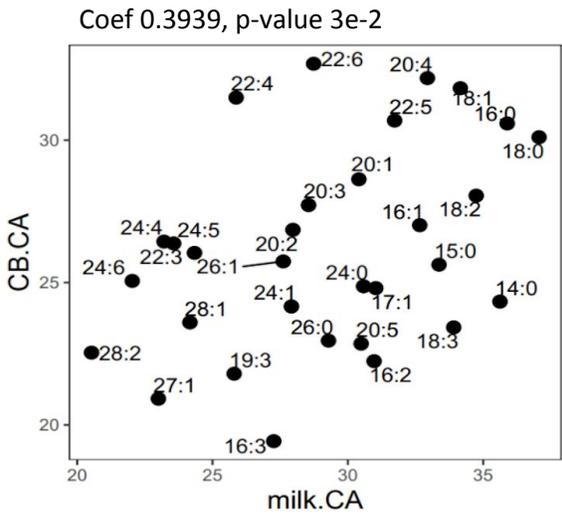
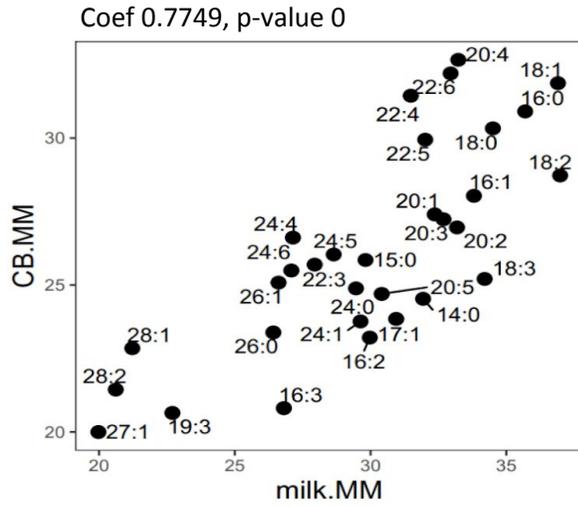
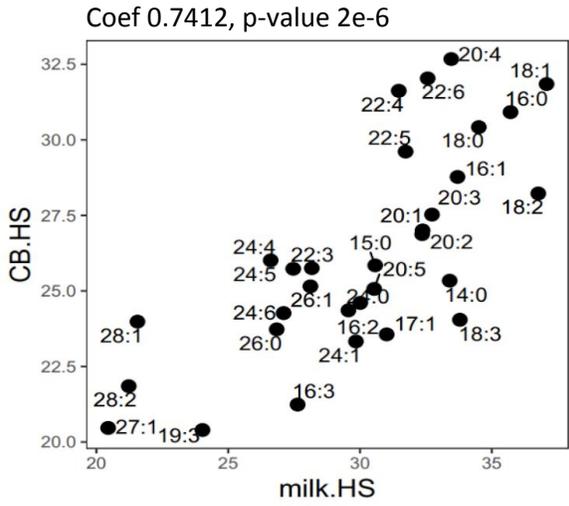


Fig. 16. Correlation between signal intensities in milk and in the brain. (A) Log-transformed upper-quartile normalized signal intensities of 31 FAs in each species: HS - humans, MM - rhesus monkeys, CA - goats and SS - pigs. The x-axis represents signal intensities in milk, the y-axis represents signal intensities in the prefrontal cortex of the brain. Pearson correlation coefficients with the p-values indicated on top. (B) Log-transformed upper-quartile normalized signal intensities of 31 FAs in the same four species. The x-axis represents signal intensities in milk, the y-axis represents signal intensities in the cerebellar grey matter. Pearson correlation coefficients with the p-values indicated on top.

To compare the relationship between the milk and brain FA composition in different species, we calculated fold-change differences, or the ratio of signal intensities, in milk and in the brain for each pair of species. We only considered FAs that were present in both milk dataset and brain dataset ($n = 31$), and we only considered species that are present in both datasets (human, macaque, pig and goat). In the plots below, for comparison between human samples and macaque samples, the \log_2FC_{milk} values were calculated as: the logarithm of the ratio of FA signal intensity in human milk to FA signal intensity in macaque milk, and the \log_2FC_{brain} was calculated as: the logarithm of the ratio of FA signal intensity in human brain to FA signal intensity in macaque brain. Pearson correlation coefficient and significance of correlation is indicated at the top of the plot (*fig. 17*). Separately, we applied t-test to compare signal intensities for each FA in human samples versus macaque samples. This analysis demonstrated that linoleic (18:2), alpha-linolenic (18:3), docosahexaenoic (22:6) and tetracosapentaenoic (24:5) FAs were present at significantly higher levels in human breast milk, as well as prefrontal cortex of the human brain (*fig. 17 A*). In cerebellum, FAs with human-specific intensity levels were: linoleic (18:2) and alpha-linoleic (18:3) (*fig. 17 B*).

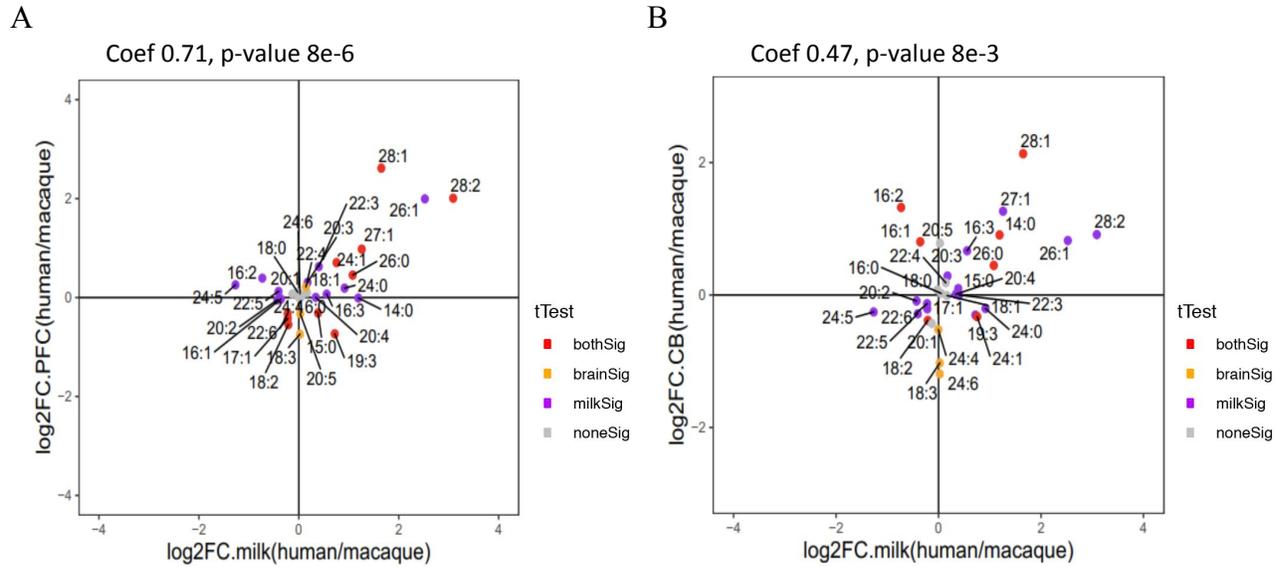


Fig. 17. Comparison of FA intensities in milk and in the brain in different species. (A) Correlation between FA signal intensities in milk and in PFC of the brain. Pearson's correlation coefficient and p-value indicated on top. Color of the FAs stands for significance of t-test between human and macaque intensities: red if significant in milk, orange if significant in brain, purple if significant in milk and in brain. The x-axis represents the ratio of mean intensity in human milk to the mean intensity in macaque milk. The y-axis represents the ratio of mean intensity in the human brain to the mean intensity in the macaque brain. (B) Correlation between FA signal intensities in milk and in CB of the brain. Pearson's correlation coefficient and p-value indicated on top. Color of the FAs stands for significance of t-test between human and macaque intensities: red if significant in milk, orange if significant in brain, purple if significant in milk and in brain. The x-axis represents the ratio of mean intensity in human milk to the mean intensity in macaque milk. The y-axis represents the ratio of mean intensity in the human brain to the mean intensity in macaque brain.

One of the most important factors that influences the composition of milk is lactation stage, or the number of days passed since childbirth. Different FAs show different intensity profiles, that can be represented as the slope of change across lactation (*fig. 18 A*). Most short and middle-chain FAs have positive slope, which means that their intensities increase with

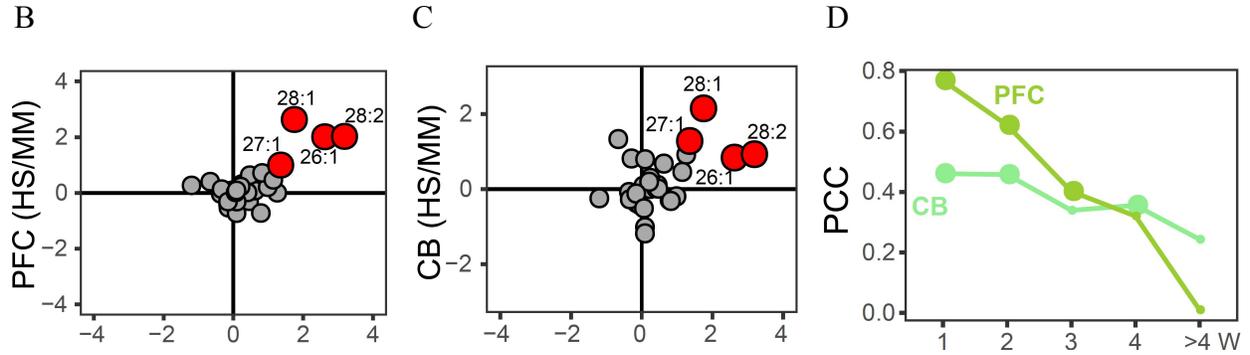
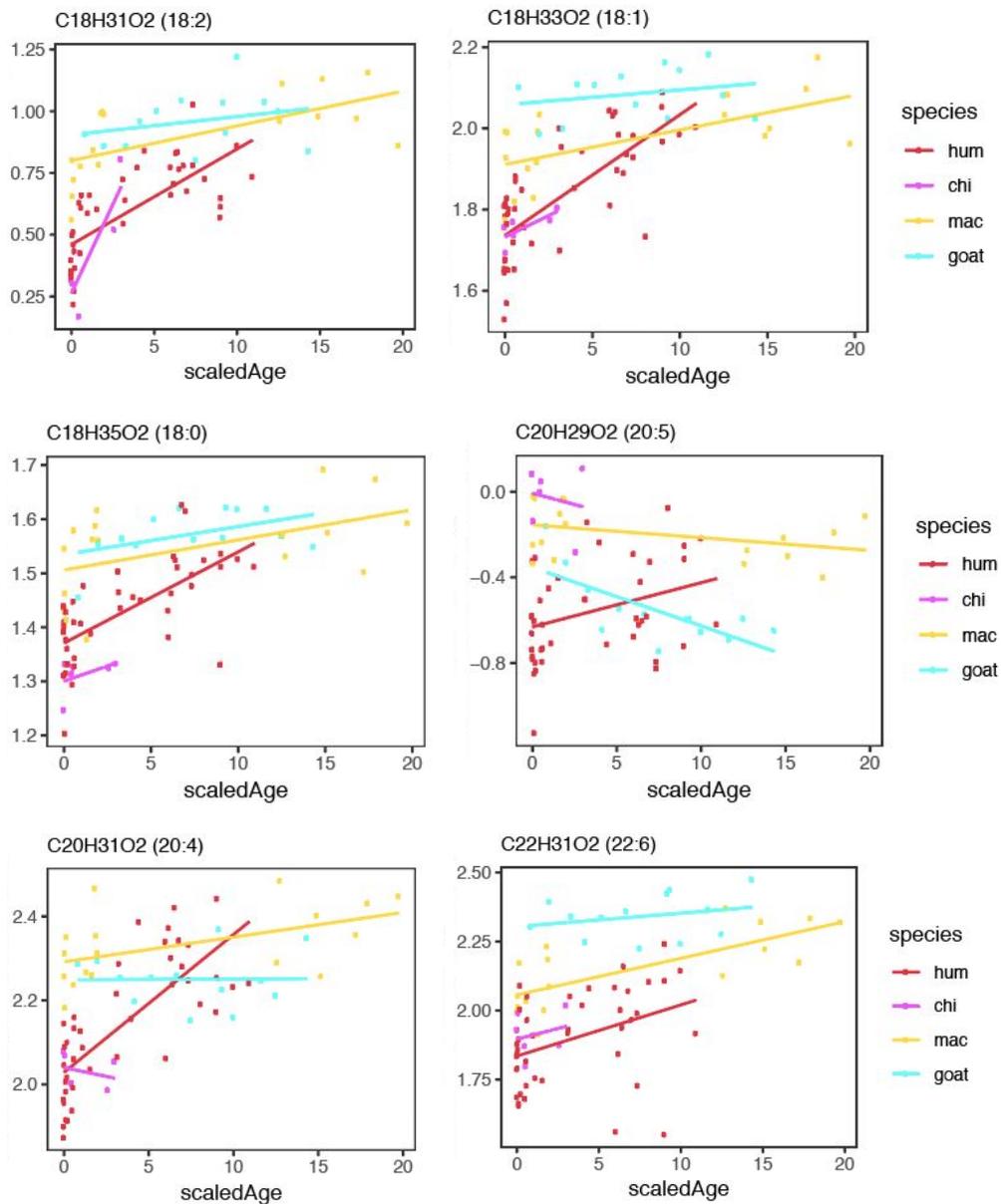


Fig. 18. Comparison of FA intensities in milk and in the brain. (A) Slope values for log-transformed, upper-quartile normalized intensities of 81 FA present in milk across \log_2 of lactation days for the human samples. FAs 26:1, 27:1, 28:1 and 28:2 are colored in red. The 95% confidence intervals are shown. (B) Correlation between FA signal intensities in milk and in PFC of the brain. The x-axis represents the fold-change difference in milk is calculated as the ratio of FA signal intensities in human milk to the ratio of FA signal intensities in macaque milk, for the first two weeks of lactation. The y-axis represents the fold-change difference in the brain is calculated as the ratio of FA signal intensities in the human prefrontal cortex to the ratio of FA signal intensities in the macaque prefrontal cortex. (C) Same as B for the cerebellum. (D) Pearson correlation coefficients calculated between FA signal intensities in milk and in the brain within lactation groups. The x-axis represents lactation weeks. The y-axis represents correlation coefficients. Larger dots correspond to p -value < 0.05 .

Similarly, signal intensities of the FAs in the human brain do not remain constant. Intensity changes across the development can be described as a slope of change which is positive for the FAs that accumulate and negative for the FAs that decrease with age. In order to assess the dynamics of FA intensity levels in the brain of different species during development, we calculated scaled age for each species multiplied by the age coefficient. Age coefficient was defined as: human maximum age divided by species maximum age, reported in (Fushan, 2015). The rounded coefficients were as follows: human - 1; chimpanzee - 2; macaque - 2,5; and goat - 5. For scaled ages we plotted concentrations for each FA separately (*fig. 19*). Most FAs tend to

increase in intensity with age: it is true for common FAs, such as stearic (18:0), oleic (18:1), linoleic (18:2) as well as docosahexaenoic (22:6) and docosapentaenoic (22:5). In some FAs, the age dynamics in humans was more pronounced than in other species: for example, in omega-6 FAs, such as arachidonic (20:4), eicosapentaenoic (20:5) and docosatetraenoic (22:4), and in some low abundant species such as 26:1, 27:1, 28:2 and 28:1. For age dependent intensity changes of all FAs in the brain please refer to supplementary *fig. S2* and *fig. S3*.



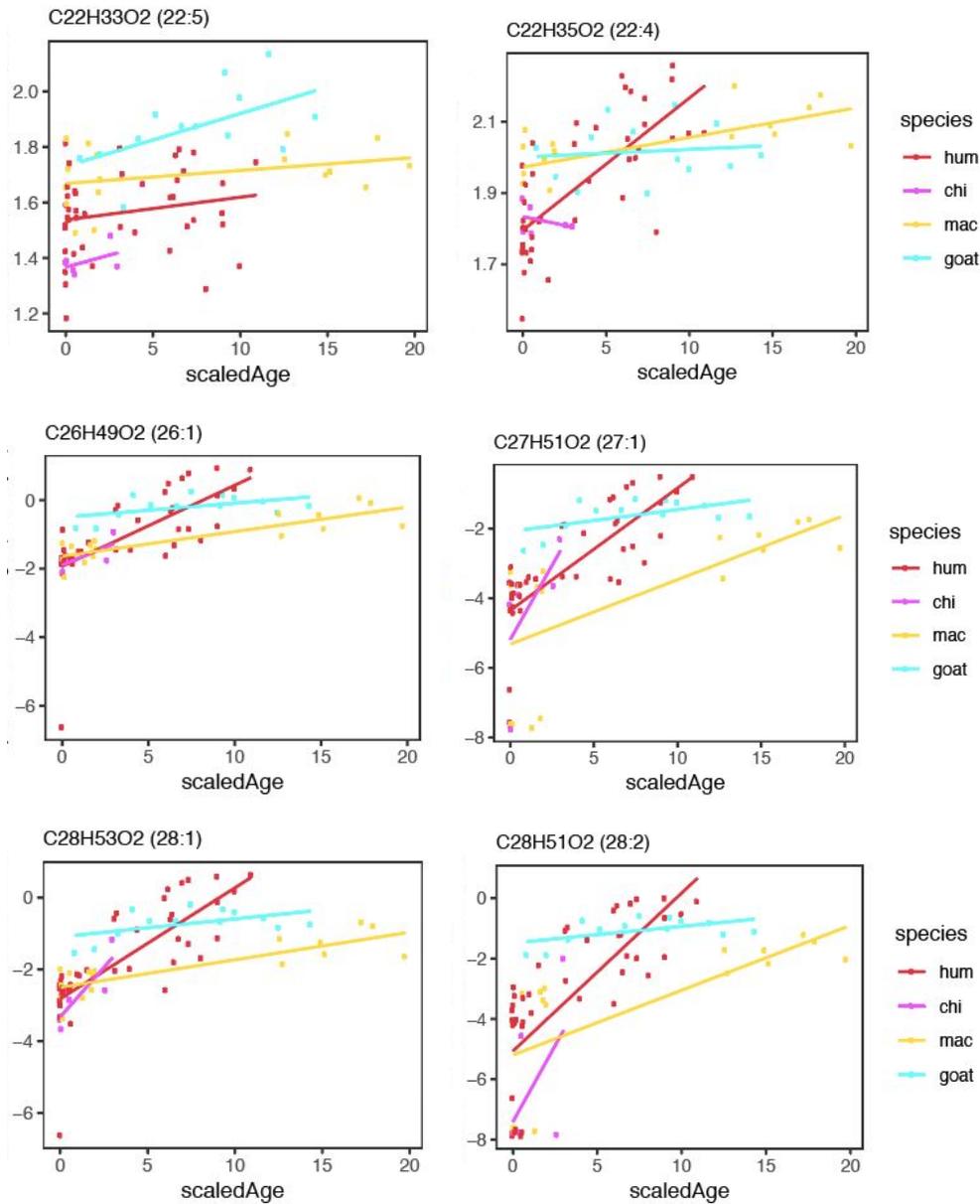


Fig. 19. Distribution of FA intensities in PFC from the age. Color represents species. The x-axis represents scaled age in days. The y-axis represents upper-quartile normalized log-transformed signal intensity. Formula of the FA is specified on top.

Using the data on the FA age dynamics in the brain, we assessed the relationship between the milk and brain FA composition in different species as the ratio of signal intensities in milk and the slope of the signal intensity changes across age in the brain. In the plots below, for

comparison between human samples and macaque samples, the \log_2FC_{milk} is calculated as the logarithm of the ratio of FA signal intensity in human milk to FA signal intensity in macaque milk, and the $slope$ is calculated as the slope of the signal intensity changes across age of the FA in human brain minus the slope of the signal intensity changes across age of the FA in macaque brain. Pearson correlation coefficient and significance of correlation are indicated at the top of the plot (*fig. 20*).

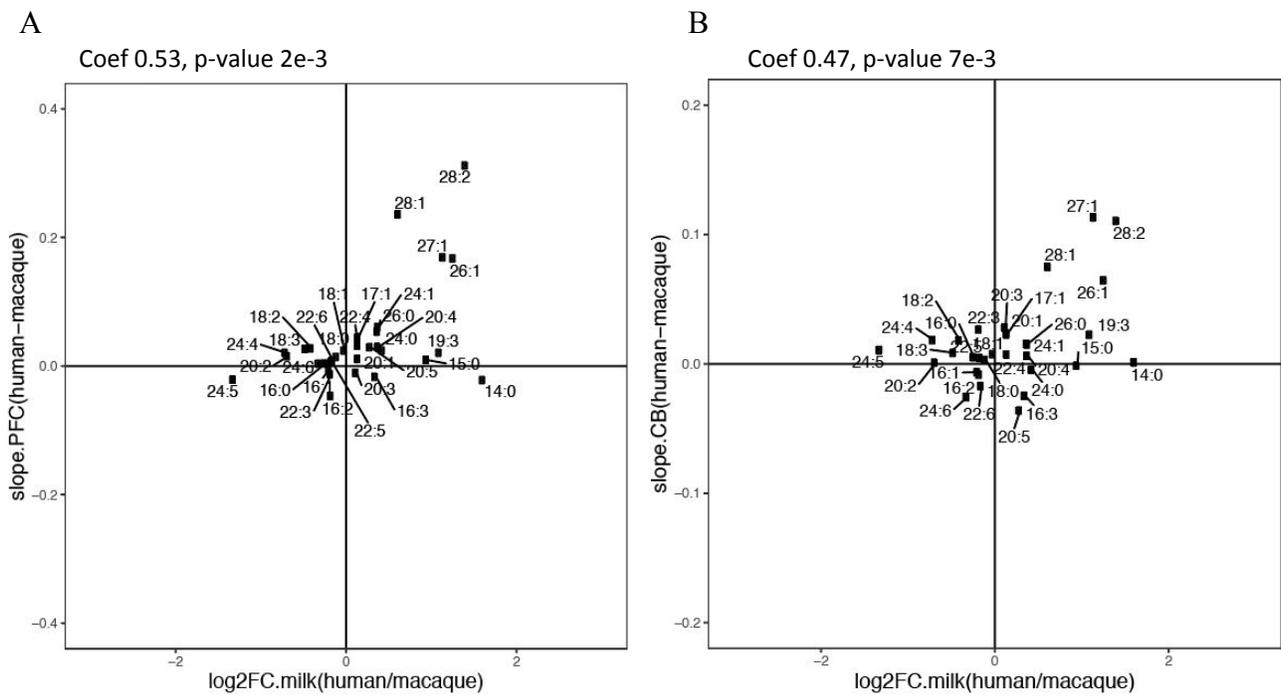


Fig. 20. Comparison of FA intensities in milk and intensity changes across age in the brain. (A) Correlation between FA signal intensities in milk and the slope of the signal intensity changes across age in the PFC of the brain. Pearson’s correlation coefficient and p-value indicated on top. The x-axis represents the ratio of mean intensity in human milk to the mean intensity in macaque milk. The y-axis represents the subtraction of the slope of the signal intensity changes across age in the macaque brain from the slope of the signal intensity changes across age in the human brain. (B) Correlation between FA signal intensities in milk and the slope of the signal intensity changes across age in CB of the brain. Pearson’s correlation coefficient and p-value indicated on

top. The x-axis represents the ratio of mean intensity in human milk to the mean intensity in macaque milk. The y-axis represents the subtraction of the slope of the signal intensity changes across age in the macaque brain from the slope of the signal intensity changes across age in the human brain.

Chapter 4. Discussion

Our study is the first comprehensive description of breast milk lipid composition as it is, and in correlation with the lipid composition of the postnatal brain of several mammalian species, including primates and in particular, humans. Here we investigate the composition of breast milk with the emphasis on triacylglycerols, the most abundant lipid class in milk, and try to trace the evolution of milk composition on 76 triacylglycerols across 7 different mammalian species: humans, rhesus monkeys, crab-eating monkeys, cows, domestic yaks, goats and pigs.

We demonstrate that bovid milk is close to primate milk in terms of middle- and long-chain saturated, unsaturated, and poly-unsaturated FAs (*fig.2B* and *fig.2C*, cluster 4; *fig. 3A*). We compare annotated TAG species between humans and three other families: bovids, pigs, and monkeys, and show that after human-monkey divergence (26 mya), humans started to produce milk with higher abundance of TAGs containing long-chain FAs with high level of desaturation (*fig. 4A*). It might be that these long-chain poly-unsaturated FAs play an important role in the development of human-specific metabolic features. Comparison of the human milk with bovid milk shows that there are mostly saturated FAs containing TAGs specific for bovids, while long-chain poly-unsaturated FAs with cumulative number of double bonds ranging from 3 to 8 have significantly higher abundance in human milk (*fig. 4B*). This demonstrates the unique characteristics of human milk lipids and impugns the appropriateness to substitute human breast milk with bovid-milk based analogs.

Comparison of the human milk with pigs milk shows changes in the opposite direction - we see that since the divergence from the last common ancestor (97 mya), pigs produced milk with higher abundance of TAGs containing very long-chain poly-unsaturated FAs. Pigs milk lipids behave similar to human milk lipids when comparing all species together (*fig. 2B* and *fig. 2C*, clusters 2 and 3) but comparison of human lipids to pig lipids reveals higher abundance of long-chain poly-unsaturated FAs in pigs milk (*fig. 4C*). One of the explanations could be shortening of the lactation period in suidae lineage and thus production of more dense milk, as opposed to humans with the extended lactation period.

We do not only check the total lipidome composition, which is mostly represented by TAGs in milk (Wastra, 1999), but we also hydrolyse milk lipids and release FAs. Principal component analysis of the milk FAs shows that hydrolyzed data is in good conformity with total lipidome data (*fig. 5* and *fig. 6*). Together with the most abundant FAs such as palmitic (16:0), stearic (18:0), oleic (18:1), linoleic (18:2) and alpha-linolenic (18:3), we also study species that constitute the minority of the FA pool, are present in only trace amounts and thus are hardly ever reported. In our data we are able to detect FAs of the chain length of up to 30 carbons (*fig. 7*).

As we shift from total lipidome in milk to studying the FA composition, we also shift from studying total lipidome of the postnatal brain, which is mostly represented by phospholipids, ceramides and sphingolipids, to studying the FAs that comprise those classes. This facilitates comparison of lipid species in milk and in the brain within one lipid class. In brain, in addition to the ones mentioned for milk, we also see high levels of arachidonic (20:4*n*-6), the most abundant omega-6 FA in the brain, and docosahexaenoic (22:6*n*-3), the most

abundant omega-3 FA, as well as their precursors and derivatives: docosapentaenoic (22:5n-3) - DHA precursor; docosatetraenoic (22:4n-6) and docosapentaenoic (22:5n-6) - AA derivatives.

When we want to study the effect of breast milk lipid composition on the lipid composition of the brain, we need to distinguish between those FAs that are synthesised endogenously from the ones that come from food. In the case of AA (20:4) and DHA (22:6) it would be difficult to execute, as these FAs show high variability in breast milk and are sensitive to the mother's diet (Martin, 2012). However, this effect is partially overcome by the study design: in our data we have representatives of two populations for the human cohort: Russian and Chinese. The population is the main factor that contributes to the differences in breast milk lipid composition, lactation stage has the second strong effect, followed by parity, mode of delivery (natural birth or Caesarean section), and baby's sex; the effect of all these factors is still minor if we compare between the species. We can assume that population factor, that accounts for genetic background and metabolic rates, is also affected by the traditional diet and eating habits that are widespread in the population. Thus, consideration of two distinct populations will diminish the effect of diet-induced variability in the FA levels.

In the brain data, we made an attempt to annotate and describe FAs of the total chain length up to 28 carbons, which is also rarely reported. Such FAs are present in trace amounts in mammals, and not in all tissues. Being part of very-long chain FAs (VLCFAs), some species of the chain length ≥ 26 are also sub-categorised as ultra long chain FAs (ULCFAs) and are only present in skin, retina, brain, testis and meibomian gland (Sassa, 2014). Meibomian gland - is a holocrine type exocrine gland, located along the edge inside the eyelid; their secret - meibum - mainly consists of lipids and prevents evaporation of the tear film from the eyes. In the brain,

ULCPUFAs are found in the sn1 position of phospholipids (Poulos, 1988). As demonstrated in rats, levels of ULCPUFAs containing phospholipids is higher in neonates than in adults, which points to their role in postnatal brain development (Robinson, 1990).

ULCFAs are synthesised by the special family of proteins called ELOVL (ELongation Of Very Long chain fatty acids), that consists of 7 members. Interestingly, each of them demonstrates substrate specificity and is able to perform synthesis only with the FAs of the particular carbon chain length and certain number of double bonds. For example, ELOVL1 elongates saturated and mono-unsaturated FAs of chain length from 20 to 26 carbons (Ohno, 2010); its 24:0 and 24:1 products are essential for the synthesis of 24-carbons long sphingolipids; *Elovl1* KO mice show severe deficiency of 24-carbon long sphingolipids (Sassa 2013). ELOVL2 elongates 20-22-carbons long polyunsaturated FAs of both n-3 and n-6 species; *Elovl2* KO mice demonstrate high levels of DPA (22:5n-6) and DHA (22:6n-3) precursors, 22:4n-6 and 22:5n-3 respectively. *Elovl2* deficiency causes near absence of ULCPUFAs (28:5n-6 and 30:5n-6) in testis with a complete arrest of spermatogenesis (Zadravec, 2011). ELOVL5 elongates 18-20-carbons long polyunsaturated FAs of both n-3 and n-6 species (Ohno, 2010); study of *Elovl5* KO mice proved its importance for LCPUFA synthesis in the liver (Moon, 2009). ELOVL3 and ELOVL7 elongate saturated and unsaturated FAs of 16-22-carbons long chain (Ohno, 2010, Naganuma 2011). *Elovl3* is expressed in sebaceous glands of the skin, hair follicles, and brown adipose tissue.

Assuming that mammary gland evolutionary developed as a hypertrophied cutaneous gland, and that the secret of the mammary gland - breast milk - contains lipids (as does the secret of sebaceous gland - sebum and the secret of the meibomian gland - meibum) there is no wonder

that some members of ELOVL family are expressed in the mammary gland, making milk a good source of the VLCFAs. So far, expression of ELOVL family members in the mammary gland was studied in several mammalian species: bovine (*Elovl1*, *Elovl5* and *Elovl2*) (Guo, 2016), rats (*Elovl5* and *Elovl4*) (Bautista, 2013), goat (*Elovl6*) (Shi, 2017; 2019) as well as in cell lines (*Elovl5* and *Elovl4*) (Mida, 2012).

In our study we show that the FAs that can be classified as ULCFAs 26:1 and 27:1 show human-specificity in milk (*fig. 17 A and B*, and *fig. 18 B and C*), and that together with FAs 28:1 and 28:2 they have higher intensity levels in human milk and the human brain. In the suggested classification for the brain FAs (saturated, even-chain unsaturated, odd-chain unsaturated and very long-chain unsaturated) they all appear in the same last group. Several long-chain poly-unsaturated FAs show human-specific intensity levels in prefrontal cortex of the human brain, these are linoleic (18:2), alpha-linolenic (18:3), docosahexaenoic (22:6) and tetracosapentaenoic (24:5), with 18:2 and 18:3 being significantly higher in cerebellum. Human breast milk also shows significantly high intensity levels of very long-chain polyunsaturated FAs: tetracosatetraenoic (24:4) and tetracosapentaenoic (24:5). Overall for the FAs that are present in both datasets (n = 31), intensities in milk correlate well with the intensities in prefrontal cortex (correlation coefficient = 0.71, p-value = 8e-6) and in cerebellum (correlation coefficient = 0.47, p-value = 8e-3) (*fig. 17*). Intensities in milk also correlate well with the intensity changes across age in prefrontal cortex (correlation coefficient = 0.53, p-value = 2e-3) and in cerebellum (correlation coefficient = 0.47, p-value = 7e-3) (*fig. 20*).

Our next step will be to study the total lipidome of the brain and define lipid classes that are specifically enriched in VLCFAs that correlate between milk and the brain (26:1, 27:1, 28:1

and 28:2). In order to clarify the human-specific interrelations between the milk lipids and brain lipids, maybe we should apply more sophisticated classification: not only consider the chain length and number of double bonds, but also consider substrate specificity of the ELOVL family members. Once the lipid classes are defined, we can then hypothesize the physical properties that these lipids possess, and their effect on the development and functioning of the nervous system.

It will also be interesting to estimate the diet effect on the accumulation of the particular brain-specific PUFAs such as arachidonic (20:4 n -6) and docosahexaenoic (22:6 n -3) FAs. Usually in animal studies the dependence between the concentrations of DHA in the diet and in the brain is assessed through feeding the newborn pups with DHA-supplemented/deficient versus regular diet. We suggest that a more proper way to study the diet effect is to provide mothers with a DHA-supplemented diet, and register DHA concentrations in mother's blood, pup's blood and brain. We could also try to assess DHA levels that are synthesised endogenously by comparing the concentration levels of its precursors, such as alpha-linolenic (18:3 n -3) and docosapentaenoic (22:5 n -3), however we will need to use more sophisticated mass-spectrometry methods to distinguish between isomers, or the n-3 and n-6 series of DPA (22:5).

Chapter 5. Conclusions

In our study we describe breast milk lipid composition and trace milk lipidome evolution in conjunction with the developing brain. We reveal substantial differences in TAG composition among seven mammalian species: three primates, three bovids, and pigs. While for most species, changes in milk lipidome composition fit the general evolutionary pattern, with distances proportional to the phylogenetic times, there is an exception. Specifically, pig milk stands out by containing unusually high amounts of long-chain polyunsaturated fatty acids. Notably, human milk is second in terms of long-chain polyunsaturated fatty acids abundance, followed by two macaque species, and then by the bovids.

Data on the FA composition of milk is in good conformity with total lipidome data. Together with the most abundant FAs: palmitic (16:0), stearic (18:0), oleic (18:1), linoleic (18:2) and alpha-linolenic (18:3), we were able to detect low abundant and thus rarely reported species with the chain length of up to 30 carbons. In the brain, we describe FAs of the total chain length up to 28 carbons, which is also rarely reported. Such FAs are present in trace amounts in mammals, and are classified as very-long chain FAs (VLCFAs) and ultra long chain FAs (ULCFAs). In our study we show that ULCFAs: 26:1 and 27:1 show human-specificity in milk, and that together with FAs 28:1 and 28:2 they have higher intensity levels in human milk and the human brain. In the suggested classification for the brain FAs (saturated, even-chain unsaturated, odd-chain unsaturated and very long-chain unsaturated) they all appear in the same last group.

Overall for the FAs that are present in both datasets ($n = 31$), intensities in milk correlate well with the intensities both in prefrontal cortex (correlation coefficient = 0.71, p -value = $8e-6$) and in cerebellum (correlation coefficient = 0.47, p -value = $8e-3$). Intensities in milk also correlate well with the intensity changes across age in prefrontal cortex (correlation coefficient = 0.53, p -value = 0.002) and in cerebellum (correlation coefficient = 0.47, p -value = $7e-3$).

Our results show that human breast milk has a unique composition of glycerolipids that is distinct from that of bovids and pigs, and most importantly even from that of primates and that lipids in milk are correlated with lipids in the brain at the level of FAs. This indicates the recent evolution in the lipid composition of breast milk and suggests an important role of breastfeeding as a way to provide essential long-chain and ultra long-chain saturated and unsaturated FAs for the proper brain development during infancy.

Bibliography

Aidoud N et al. 2018. A combination of lipidomics, MS imaging, and PET scan imaging reveals differences in cerebral activity in rat pups according to the lipid quality of infant formulas. *FASEB J.* 32:4776–4790.

Ali A et al. 2019. Developmental Vitamin D Deficiency Produces Behavioral Phenotypes of Relevance to Autism in an Animal Model. *Nutrients.* 11. doi: 10.3390/nu11051187.

Andreas NJ et al. 2016. Impact of maternal BMI and sampling strategy on the concentration of leptin, insulin, ghrelin and resistin in breast milk across a single feed: a longitudinal cohort study. *BMJ Open.* 6:e010778. doi: 10.1136/bmjopen-2015-010778.

Anwar MA et al. 2015. Optimization of metabolite extraction of human vein tissue for ultra performance liquid chromatography-mass spectrometry and nuclear magnetic resonance-based untargeted metabolic profiling. *The Analyst.* 140:7586–7597. doi: 10.1039/c5an01041a.

Baitchman EJ, Tlusty MF, Murphy HW. 2007. Passive transfer of maternal antibodies to West Nile virus in flamingo chicks (*Phoenicopterus chilensis* and *Phoenicopterus ruber ruber*). *J. Zoo Wildl. Med.* 38:337–340.

Ballard O, Morrow AL. 2013. Human Milk Composition. *Pediatric Clinics of North America*. 60:49–74. doi: 10.1016/j.pcl.2012.10.002.

Bautista CJ et al. 2013. Protein restriction in the rat negatively impacts long-chain polyunsaturated fatty acid composition and mammary gland development at the end of gestation. *Arch. Med. Res.* 44:429–436.

Beck KL et al. 2015. Comparative Proteomics of Human and Macaque Milk Reveals Species-Specific Nutrition during Postnatal Development. *Journal of Proteome Research*. 14:2143–2157. doi: 10.1021/pr501243m.

Bhinder G et al. 2017. Milk Fat Globule Membrane Supplementation in Formula Modulates the Neonatal Gut Microbiome and Normalizes Intestinal Development. *Sci. Rep.* 7:45274.

Blanksby SJ, Mitchell TW. 2010. Advances in mass spectrometry for lipidomics. *Annu. Rev. Anal. Chem.* 3:433–465.

Bollongino, Ruth, Joachim Burger, Adam Powell, Marjan Mashkour, Jean-Denis Vigne, and Mark G. Thomas. 2012. Modern Taurine Cattle Descended from Small Number of near-Eastern Founders. *Molecular Biology and Evolution* 29 (9): 2101–4.

Bowden JA et al. 2017. Harmonizing lipidomics: NIST interlaboratory comparison exercise for lipidomics using SRM 1950-Metabolites in Frozen Human Plasma. *J. Lipid Res.* 58:2275–2288.

Bozek K et al. 2015. Organization and evolution of brain lipidome revealed by large-scale analysis of human, chimpanzee, macaque, and mouse tissues. *Neuron.* 85:695–702.

Bromke MA et al. 2015. Liquid chromatography high-resolution mass spectrometry for fatty acid profiling. *Plant J.* 81:529–536.

Butts CA et al. 2018. Human Milk Composition and Dietary Intakes of Breastfeeding Women of Different Ethnicity from the Manawatu-Wanganui Region of New Zealand. *Nutrients.* 10. doi: 10.3390/nu10091231.

Chang N et al. 2015. Macronutrient composition of human milk from Korean mothers of full term infants born at 37-42 gestational weeks. *Nutr. Res. Pract.* 9:433–438.

Chen W et al. 2018. Lactation Stage-Dependency of the Sow Milk Microbiota. *Front. Microbiol.* 9:945.

Cohen Engler A, Hadash A, Shehadeh N, Pillar G. 2012. Breastfeeding may improve nocturnal sleep and reduce infantile colic: Potential role of breast milk melatonin. *Eur. J. Pediatr.* 171:729–732.

Consortium LSI, Lipidomics Standards Initiative Consortium. 2019. Lipidomics needs more standardization. *Nature Metabolism*. 1:745–747. doi: 10.1038/s42255-019-0094-z.

Dekaban AS. 1978. Changes in brain weights during the span of human life: relation of brain weights to body heights and body weights. *Ann. Neurol*. 4:345–356.

Demmelair H, Prell C, Timby N, Lönnerdal B. 2017. Benefits of Lactoferrin, Osteopontin and Milk Fat Globule Membranes for Infants. *Nutrients*. 9. doi: 10.3390/nu9080817.

Dunn WB et al. 2011. Procedures for large-scale metabolic profiling of serum and plasma using gas chromatography and liquid chromatography coupled to mass spectrometry. *Nature Protocols*. 6:1060–1083. doi: 10.1038/nprot.2011.335.

Eastham NT et al. 2018. Associations between age at first calving and subsequent lactation performance in UK Holstein and Holstein-Friesian dairy cows. *PLoS One*. 13:e0197764.

Eidelman AI, Schanler RJ. 2012. Breastfeeding and the use of human milk. *Pediatrics*. <https://psycnet.apa.org/record/2012-09899-031>.

Eisert R. 2011. Hypercarnivory and the brain: protein requirements of cats reconsidered. *J. Comp. Physiol. B*. 181:1–17.

Ejendal, Karin F. K., and Christine A. Hrycyna. 2002. Multidrug Resistance and Cancer: The Role of the Human ABC Transporter ABCG2. *Current Protein & Peptide Science* 3 (5): 503–11.

Else PL. 2013. Dinosaur lactation? *J. Exp. Biol.* 216:347–351.

Fahy E et al. 2005. A comprehensive classification system for lipids. *Journal of Lipid Research.* 46:839–862. doi: 10.1194/jlr.e400004-jlr200.

Fong BY, Norris CS, MacGibbon AKH. 2007. Protein and lipid composition of bovine milk-fat-globule membrane. *Int. Dairy J.* 17:275–288.

Frelinger JA. 1971. Maternally derived transferrin in pigeon squabs. *Science.* 171:1260–1261.

Fushan AA et al. 2015. Gene expression defines natural changes in mammalian lifespan. *Aging Cell.* 14:352–365.

Galante L et al. 2018. Sex-Specific Human Milk Composition: The Role of Infant Sex in Determining Early Life Nutrition. *Nutrients.* 10. doi: 10.3390/nu10091194.

Galli C, Trzeciak HI, Paoletti R. 1971. Effects of dietary fatty acids on the fatty acid composition of brain ethanolamine phosphoglyceride: Reciprocal replacement of n-6 and n-3

polyunsaturated fatty acids. *Biochimica et Biophysica Acta (BBA) - Lipids and Lipid Metabolism*. 248:449–454. doi: 10.1016/0005-2760(71)90233-5.

Cohen-Zinder, Miri, Eyal Seroussi, Denis M. Larkin, Juan J. Looor, Annelie Everts-van der Wind, Jun-Heon Lee, James K. Drackley, et al. 2005. Identification of a Missense Mutation in the Bovine ABCG2 Gene with a Major Effect on the QTL on Chromosome 6 Affecting Milk Yield and Composition in Holstein Cattle. *Genome Research* 15 (7): 936–44.

Gottesman, Michael M., Tito Fojo, and Susan E. Bates. 2002. Multidrug Resistance in Cancer: Role of ATP-Dependent Transporters. *Nature Reviews. Cancer* 2 (1): 48–58.

Goudswaard J, van der Donk JA, van der Gaag I, Noordzij A. 1979. Peculiar IgA transfer in the pigeon from mother to squab. *Dev. Comp. Immunol.* 3:307–319.

Grande Covián F. 1979. [Energy metabolism of the brain in children (author's transl)]. *An. Esp. Pediatr.* 12:235–244.

Granot E, Ishay-Gigi K, Malaach L, Flidel-Rimon O. 2016. Is there a difference in breast milk fatty acid composition of mothers of preterm and term infants? *J. Matern. Fetal. Neonatal Med.* 29:832–835.

Grisart, Bernard, Wouter Coppieters, Frédéric Farnir, Latifa Karim, Christine Ford, Paulette Berzi, Nadine Cambisano, et al. 2002. Positional Candidate Cloning of a QTL in Dairy Cattle: Identification of a Missense Mutation in the Bovine DGAT1 Gene with Major Effect on Milk Yield and Composition. *Genome Research* 12 (2): 222–31.

Guan, Z. Z., M. Söderberg, P. Sindelar, and C. Edlund. 1994. Content and Fatty Acid Composition of Cardiolipin in the Brain of Patients with Alzheimer's Disease. *Neurochemistry International* 25 (3): 295–300.

Guo Z et al. 2016. Rapamycin Inhibits Expression of Elongation of Very-long-chain Fatty Acids 1 and Synthesis of Docosahexaenoic Acid in Bovine Mammary Epithelial Cells. *Asian-australas. J. Anim. Sci.* 29:1646–1652.

Hampel D, Shahab-Ferdows S, Islam MM, Peerson JM, Allen LH. 2017. Vitamin Concentrations in Human Milk Vary with Time within Feed, Circadian Rhythm, and Single-Dose Supplementation. *J. Nutr.* 147:603–611.

Hinde K et al. 2013. Daughter dearest: Sex-biased calcium in mother's milk among rhesus macaques. *American Journal of Physical Anthropology.* 151:144–150. doi: 10.1002/ajpa.22229.

Hinde K. 2009. Richer milk for sons but more milk for daughters: Sex-biased investment during lactation varies with maternal life history in rhesus macaques. *Am. J. Hum. Biol.* <https://onlinelibrary.wiley.com/doi/abs/10.1002/ajhb.20917>.

Hinde K, Milligan LA. 2011. Primate milk: proximate mechanisms and ultimate perspectives. *Evol. Anthropol.* 20:9–23.

Innis SM. 2007. Dietary (n-3) Fatty Acids and Brain Development. *The Journal of Nutrition.* 137:855–859. doi: 10.1093/jn/137.4.855.

Innis SM. 1992. Human milk and formula fatty acids. *J. Pediatr.* 120:S56–61.

Innis SM. 2014. Impact of maternal diet on human milk composition and neurological development of infants. *Am. J. Clin. Nutr.* 99:734S–41S.

Isaacs EB et al. 2010. Impact of breast milk on intelligence quotient, brain size, and white matter development. *Pediatr. Res.* 67:357–362.

Jiang J et al. 2016. Changes in fatty acid composition of human milk over lactation stages and relationship with dietary intake in Chinese women. *Food Funct.* 7:3154–3162.

Jonker, Johan W., Gracia Merino, Sandra Musters, Antonius E. van Herwaarden, Ellen Bolscher, Els Wagenaar, Elly Mesman, Trevor C. Dale, and Alfred H. Schinkel. 2005. The Breast Cancer Resistance Protein BCRP (ABCG2) Concentrates Drugs and Carcinogenic Xenotoxins into Milk. *Nature Medicine* 11 (2): 127–29.

Knickmeyer RC et al. 2008. A structural MRI study of human brain development from birth to 2 years. *J. Neurosci.* 28:12176–12182.

Kuhl C, Tautenhahn R, Böttcher C, Larson TR, Neumann S. 2012. CAMERA: an integrated strategy for compound spectra extraction and annotation of liquid chromatography/mass spectrometry data sets. *Anal. Chem.* 84:283–289.

Lee H, Zavaleta N, Chen S-Y, Lönnerdal B, Slupsky C. 2018. Effect of bovine milk fat globule membranes as a complementary food on the serum metabolome and immune markers of 6-11-month-old Peruvian infants. *npj Science of Food*. 2. doi: 10.1038/s41538-018-0014-8.

Lenz EM, Wilson ID. 2007. Analytical strategies in metabonomics. *J. Proteome Res.* 6:443–458.

Levitt P. 2003. Structural and functional maturation of the developing primate brain. *J. Pediatr.* 143:S35–45.

Li Q et al. 2017. Changes in Lipidome Composition during Brain Development in Humans, Chimpanzees, and Macaque Monkeys. *Mol. Biol. Evol.* 34:1155–1166.

Libiseller G et al. 2015. IPO: a tool for automated optimization of XCMS parameters. *BMC Bioinformatics.* 16:118.

Liu Z, Cocks BG, Rochfort S. 2016. Comparison of Molecular Species Distribution of DHA-Containing Triacylglycerols in Milk and Different Infant Formulas by Liquid Chromatography–Mass Spectrometry. *J. Agric. Food Chem.* 64:2134–2144.

Lopez C, Madec M-N, Jimenez-Flores R. 2010. Lipid rafts in the bovine milk fat globule membrane revealed by the lateral segregation of phospholipids and heterogeneous distribution of glycoproteins. *Food Chem.* 120:22–33.

Lopez C, Ménard O. 2011. Human milk fat globules: Polar lipid composition and in situ structural investigations revealing the heterogeneous distribution of proteins and the lateral segregation of sphingomyelin in the biological membrane. *Colloids and Surfaces B: Biointerfaces.* 83:29–41. doi: 10.1016/j.colsurfb.2010.10.039.

Ma, Xiaoxiao, and Yu Xia. 2014. Pinpointing Double Bonds in Lipids by Paternò-Büchi Reactions and Mass Spectrometry. *Angewandte Chemie, International Edition* 53 (10): 2592–96.

Martin MA et al. 2012. Fatty acid composition in the mature milk of Bolivian forager-horticulturalists: controlled comparisons with a US sample. *Maternal & Child Nutrition*. 8:404–418. doi: 10.1111/j.1740-8709.2012.00412.x.

McCann JC, Ames BN. 2008. Is there convincing biological or behavioral evidence linking vitamin D deficiency to brain dysfunction? *FASEB J*. 22:982–1001.

Mida K, Shamay A, Argov-Argaman N. 2012. Elongation and desaturation pathways in mammary gland epithelial cells are associated with modulation of fat and membrane composition. *J. Agric. Food Chem*. 60:10657–10665.

Mitchell, Todd W., Huong Pham, Michael C. Thomas, and Stephen J. Blanksby. 2009. Identification of Double Bond Position in Lipids: From GC to OzID. *Journal of Chromatography. B, Analytical Technologies in the Biomedical and Life Sciences* 877 (26): 2722–35.

Moon Y-A, Hammer RE, Horton JD. 2009. Deletion of ELOVL5 leads to fatty liver through activation of SREBP-1c in mice. *J. Lipid Res*. 50:412–423.

Moore SA, Yoder E, Murphy S, Dutton GR, Spector AA. 1991. Astrocytes, not neurons, produce docosahexaenoic acid (22: 6 ω -3) and arachidonic acid (20: 4 ω -6). *J. Neurochem*. 56:518–524.

Mulder KA, Elango R, Innis SM. 2018. Fetal DHA inadequacy and the impact on child neurodevelopment: a follow-up of a randomised trial of maternal DHA supplementation in pregnancy. *Br. J. Nutr.* 119:271–279.

Naganuma T, Sato Y, Sassa T, Ohno Y, Kihara A. 2011. Biochemical characterization of the very long-chain fatty acid elongase ELOVL7. *FEBS Lett.* <https://febs.onlinelibrary.wiley.com/doi/abs/10.1016/j.febslet.2011.09.024>.

Nicholson JK, Connelly J, Lindon JC, Holmes E. 2002. Metabonomics: a platform for studying drug toxicity and gene function. *Nature Reviews Drug Discovery.* 1:153–161. doi: 10.1038/nrd728.

Nicholson JK, Lindon JC. 2008. Systems biology: Metabonomics. *Nature.* 455:1054–1056.

O'Brien, J. S., and E. L. Sampson. 1965. Lipid Composition of the Normal Human Brain: Gray Matter, White Matter, and Myelin. *Journal of Lipid Research* 6 (4): 537–44.

Oecd, OECD. 2014. Infant and young child feeding. doi: 10.1787/health_glance_ap-2014-20-en.

Oftedal OT. 2012. The evolution of milk secretion and its ancient origins. *Animal.* 6:355–368.

Oftedal OT. 2002. The mammary gland and its origin during synapsid evolution. *J. Mammary Gland Biol. Neoplasia*. 7:225–252.

Oftedal OT, Bowen WD, Boness DJ. 1993. Energy Transfer by Lactating Hooded Seals and Nutrient Deposition in Their Pups during the Four Days from Birth to Weaning. *Physiol. Zool*. 66:412–436.

Oftedal OT, Dhouailly D. 2013. Evo-devo of the mammary gland. *J. Mammary Gland Biol. Neoplasia*. 18:105–120.

Oftedal OT, Iverson SJ. 1995. Comparative Analysis of Nonhuman Milks. *Handbook of Milk Composition*. 749–789. doi: 10.1016/b978-012384430-9/50035-4.

Ohno Y et al. 2010. ELOVL1 production of C24 acyl-CoAs is linked to C24 sphingolipid synthesis. *Proc. Natl. Acad. Sci. U. S. A.* 107:18439–18444.

Panganamala, R. V., L. A. Horrocks, J. C. Geer, and D. G. Cornwell. 1971. Positions of Double Bonds in the Monounsaturated Alk-1-Enyl Groups from the Plasmalogens of Human Heart and Brain. *Chemistry and Physics of Lipids* 6 (2): 97–102.

Poulos A, Sharp P, Johnson D, Easton C. 1988. The occurrence of polyenoic very long chain fatty acids with greater than 32 carbon atoms in molecular species of phosphatidylcholine in

normal and peroxisome-deficient (Zellweger's syndrome) brain. *Biochem. J.* 253:645–650. doi: 10.1042/bj2530645.

Powe CE, Knott CD, Conklin-Brittain N. 2010. Infant sex predicts breast milk energy content. *Am. J. Hum. Biol.* 22:50–54.

Robinson BS, Johnson DW, Poulos A. 1990. Unique molecular species of phosphatidylcholine containing very-long-chain (C24-C38) polyenoic fatty acids in rat brain. *Biochem. J.* <http://www.biochemj.org/content/265/3/763.abstract>.

Rouser, G., C. Galli, and G. Kritchevsky. 1965. LIPID CLASS COMPOSITION OF NORMAL HUMAN BRAIN AND VARIATIONS IN METACHROMATIC LEUCODYSTROPHY, TAY-SACHS, NIEMANN-PICK, CHRONIC GAUCHER'S AND ALZHEIMER'S DISEASES. *Journal of the American Oil Chemists' Society* 42 (May): 404–10.

Saad K et al. 2016. Vitamin D status in autism spectrum disorders and the efficacy of vitamin D supplementation in autistic children. *Nutr. Neurosci.* 19:346–351.

Sánchez CL et al. 2009. The possible role of human milk nucleotides as sleep inducers. *Nutr. Neurosci.* 12:2–8.

Sarafian MH et al. 2014. Objective Set of Criteria for Optimization of Sample Preparation Procedures for Ultra-High Throughput Untargeted Blood Plasma Lipid Profiling by Ultra Performance Liquid Chromatography–Mass Spectrometry. *Analytical Chemistry*. 86:5766–5774. doi: 10.1021/ac500317c.

Sassa T, Kihara A. 2014. Metabolism of very long-chain Fatty acids: genes and pathophysiology. *Biomol. Ther.* 22:83–92.

Sastry, P. S. 1985. Lipids of Nervous Tissue: Composition and Metabolism. *Progress in Lipid Research* 24 (2): 69–176.

Scheu, Amelie, Adam Powell, Ruth Bollongino, Jean-Denis Vigne, Anne Tresset, Canan Çakırlar, Norbert Benecke, and Joachim Burger. 2015. The Genetic Prehistory of Domesticated Cattle from Their Origin to the Spread across Europe. *BMC Genetics* 16 (May): 54.

Schefers, Jonathan M., and Kent A. Weigel. 2012. Genomic Selection in Dairy Cattle: Integration of DNA Testing into Breeding Programs. *Animal Frontiers* 2 (1): 4–9.

Shetty S, Bharathi L, Shenoy KB, Hegde SN. 1992. Biochemical properties of pigeon milk and its effect on growth. *J. Comp. Physiol. B*. 162:632–636.

Shi H et al. 2019. Fatty Acid Elongase 7 (ELOVL7) Plays a Role in the Synthesis of Long-Chain Unsaturated Fatty Acids in Goat Mammary Epithelial Cells. *Animals (Basel)*. 9. doi: 10.3390/ani9060389.

Shi HB et al. 2017. Fatty acid elongase 6 plays a role in the synthesis of long-chain fatty acids in goat mammary epithelial cells. *J. Dairy Sci.* 100:4987–4995.

Smith A. 2000. *Oxford Dictionary of Biochemistry and Molecular Biology: Revised Edition*. Oxford University Press.

Smith CA, Want EJ, O'Maille G, Abagyan R, Siuzdak G. 2006. XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. *Anal. Chem.* 78:779–787.

Smith, S. J., S. Cases, D. R. Jensen, H. C. Chen, E. Sande, B. Tow, D. A. Sanan, J. Raber, R. H. Eckel, and R. V. Farese Jr. 2000. Obesity Resistance and Multiple Mechanisms of Triglyceride Synthesis in Mice Lacking Dgat. *Nature Genetics* 25 (1): 87–90.

Sprecher H. 2000. Metabolism of highly unsaturated n-3 and n-6 fatty acids. *Biochem Biophys Acta.* 1486:219–231.

Tanaka, Koichi, Hiroaki Waki, Yutaka Ido, Satoshi Akita, Yoshikazu Yoshida, Tamio Yoshida, and T. Matsuo. 1988. Protein and Polymer Analyses up to m/z 100 000 by Laser Ionization Time-of-Flight Mass Spectrometry. *Rapid Communications in Mass Spectrometry: RCM* 2 (August): 151.

Tau GZ, Peterson BS. 2010. Normal development of brain circuits. *Neuropsychopharmacology*. 35:147–168.

Thakkar SK et al. 2013. Dynamics of human milk nutrient composition of women from Singapore with a special focus on lipids. *Am. J. Hum. Biol.* 25:770–779.

VandeHaar, M. J., L. E. Armentano, and K. Weigel. 2016. Harnessing the Genetics of the Modern Dairy Cow to Continue Improvements in Feed Efficiency. *Journal of Dairy*. <https://www.sciencedirect.com/science/article/pii/S0022030216301655>.

Vorkas PA et al. 2015. Untargeted UPLC-MS profiling pipeline to expand tissue metabolome coverage: application to cardiovascular disease. *Anal. Chem.* 87:4184–4193.

Wäldchen, Fabian, Simon Becher, Patrick Esch, Mario Kompauer, and Sven Heiles. 2017. Selective Phosphatidylcholine Double Bond Fragmentation and Localisation Using Paternò-Büchi Reactions and Ultraviolet Photodissociation. *The Analyst* 142 (24): 4744–55.

Walfisch A, Sermer C, Cressman A, Koren G. 2013. Breast milk and cognitive development—the role of confounders: a systematic review. *BMJ Open*. 3:e003259.

Want EJ et al. 2013. Global metabolic profiling of animal and human tissues via UPLC-MS. *Nat. Protoc.* 8:17–32.

Ward RE, Niñonuevo M, Mills DA, Lebrilla CB, German JB. 2006. In vitro fermentation of breast milk oligosaccharides by *Bifidobacterium infantis* and *Lactobacillus gasseri*. *Appl. Environ. Microbiol.* 72:4497–4499.

Wenk MR. 2010. Lipidomics: new tools and applications. *Cell*. 143:888–895.

Wenk MR. 2005. The emerging field of lipidomics. *Nat. Rev. Drug Discov.* 4:594–610.

Wong BH et al. 2016. Mfsd2a Is a Transporter for the Essential ω -3 Fatty Acid Docosahexaenoic Acid (DHA) in Eye and Is Important for Photoreceptor Cell Development. *J. Biol. Chem.* 291:10501–10514.

World Health Organisation Staff, World Health Organization. 2003. *Global Strategy for Infant and Young Child Feeding*. World Health Organization.

Wu X, Jackson RT, Khan SA, Ahuja J, Pehrsson PR. 2018. Human Milk Nutrient Composition in the United States: Current Knowledge, Challenges, and Research Needs. *Curr Dev Nutr.* 2:nzy025.

Yamashita, Masamichi, and John B. Fenn. 1984. Electrospray Ion Source. Another Variation on the Free-Jet Theme. *The Journal of Physical Chemistry* 88 (20): 4451–59.

Appendices

	Total number of samples	Parity					Parturition			Baby's Sex		
		1st	2nd	3rd	4th	undef.	N	C	undef.	F	M	undef.
Moscow cohort	297	50	71	14	6	156	142	68	87	136	74	87
Shanghai cohort	291	167	91	0	0	33	106	155	30	140	126	25

Tbl. 1. Meta information on the human milk samples.

	Baby's height at birth	Baby's weight at birth	Mother's mean age	Mother's mean height	Mother's mean weight
Moscow cohort	51	3300	31	166	60
Shanghai cohort	51	3200	32	165	63

Tbl. 2. Additional information on the volunteers.

	Total number of samples	Region		Age (days)			Sex	
		PFC	CB	min	max	med	F	M
Human	92	43	49	0	365	30	52	40
Chimpanzee	18	9	9	0	45	7.5	8	10
Macaque	38	19	19	0	358	24	2	36
Pig	20	10	10	-	-	-	-	-
Goat	26	13	13	5	86	45	14	12

Tbl. 3. Meta information on the brain samples.

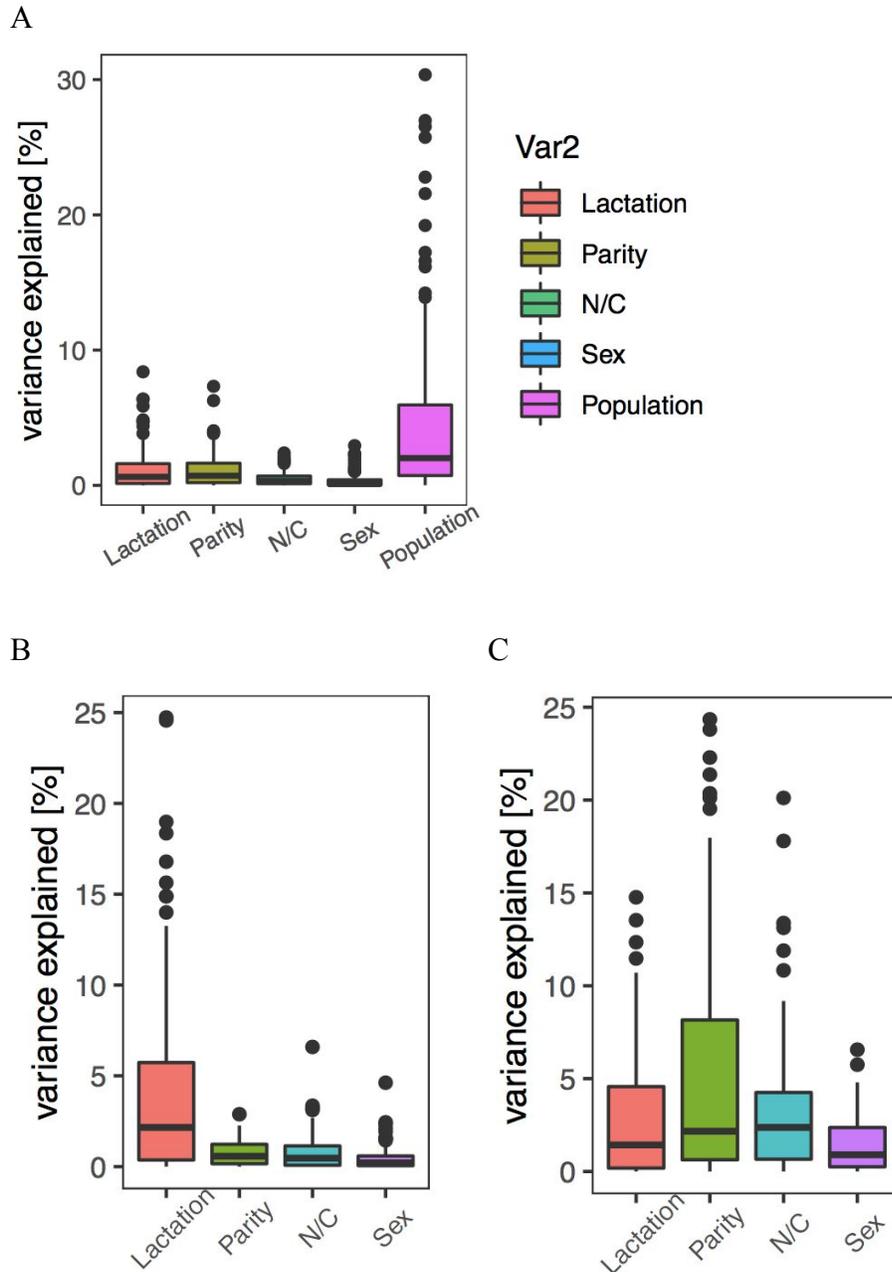
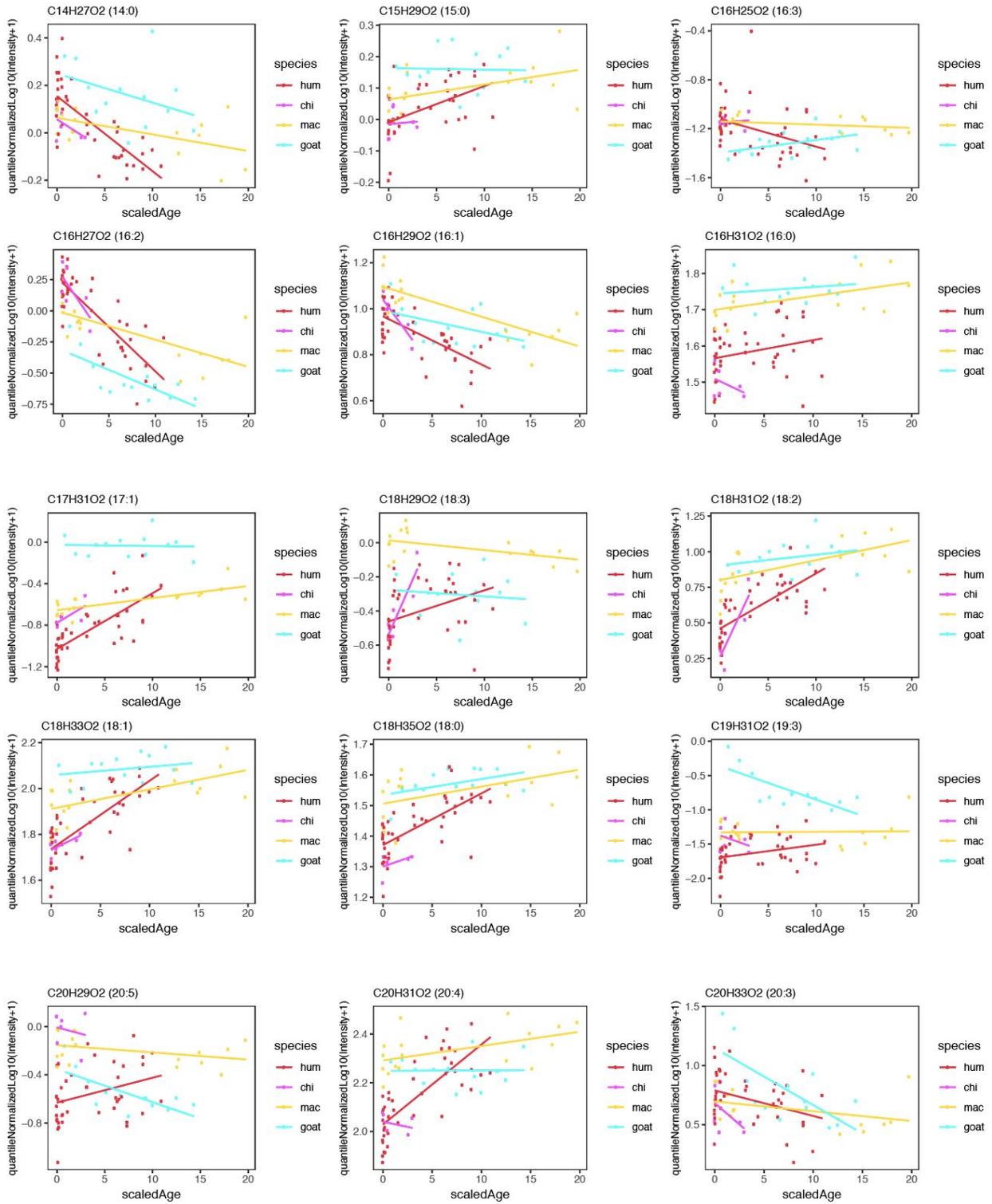
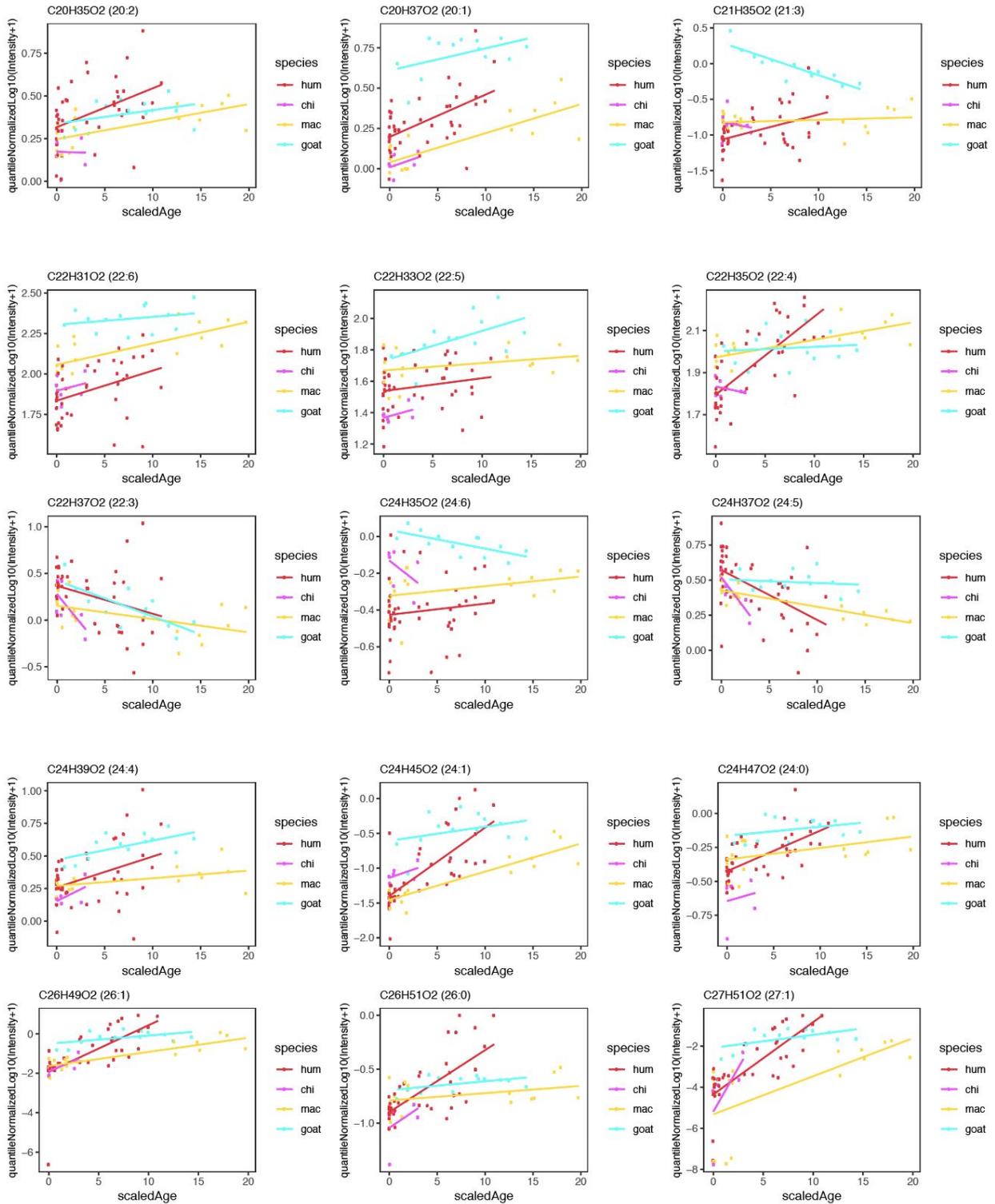


Fig. S1. Factors underlying the differences between human milk samples. (A) Variance for all human breast milk samples and main factors: Lactation (lactation stage in days), Parity (number of previously born children), N/C (mode of delivery - natural or Cesarean section), Sex (baby's sex). (B) Variance for human samples from the Shanghai group. (C) Variance for human samples from the Moscow group.





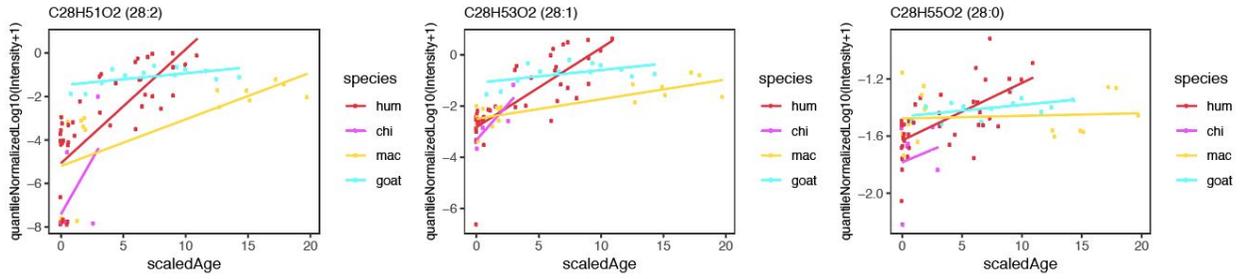
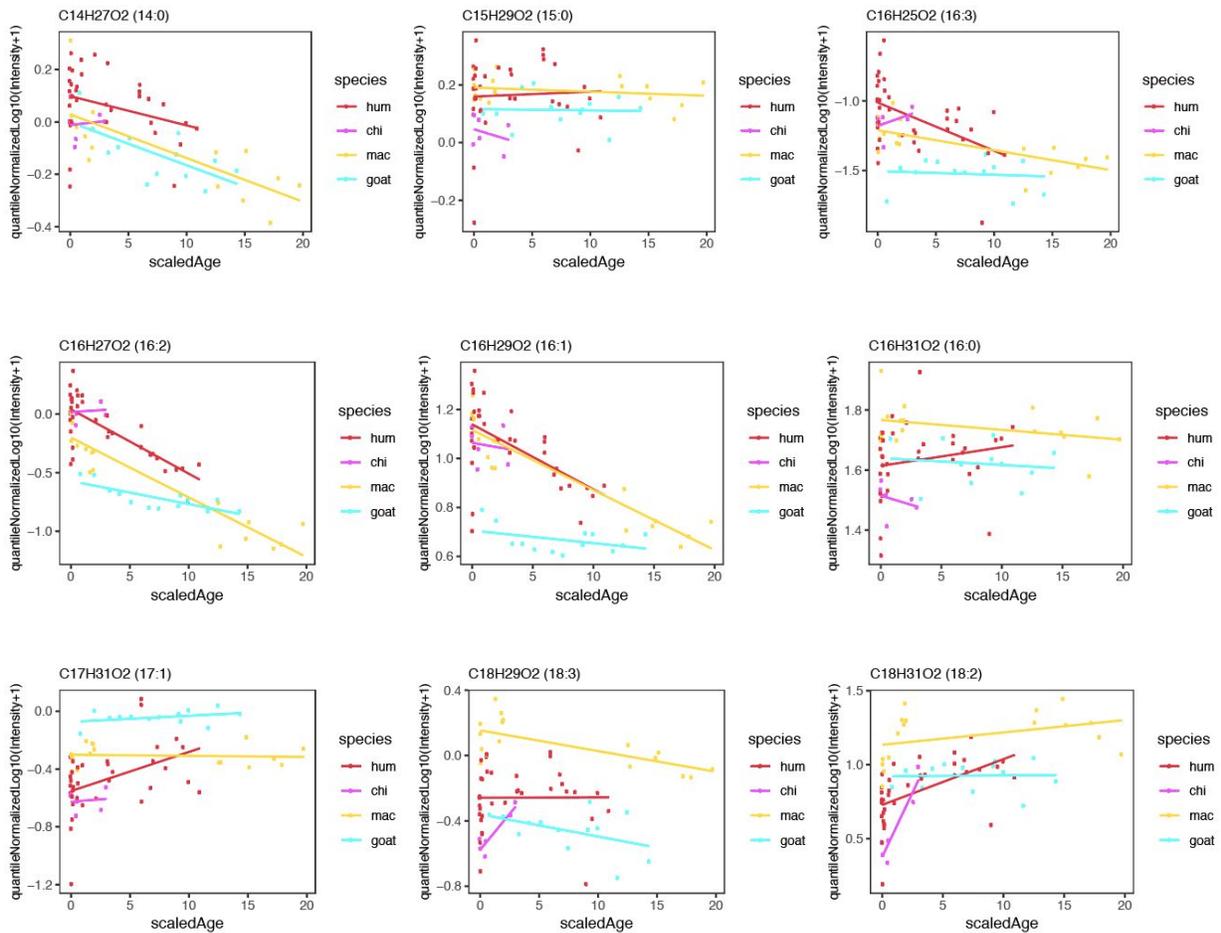
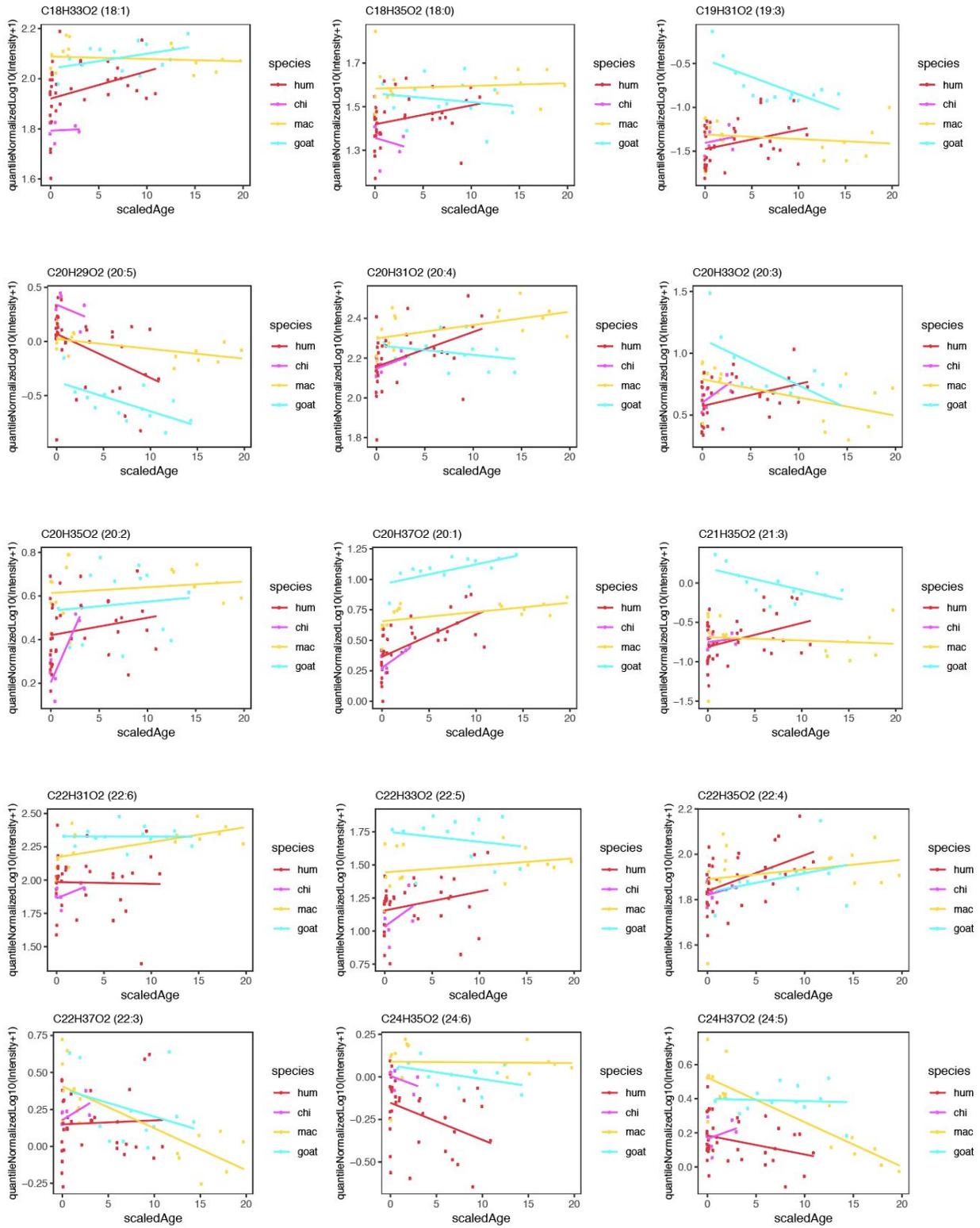


Fig. S2. Dependency of FA intensities in PFC from the age. Color represents species. The x-axis represents scaled age in days. The y-axis represents upper-quartile normalized log-transformed signal intensity. Formula of the FA is specified on top.





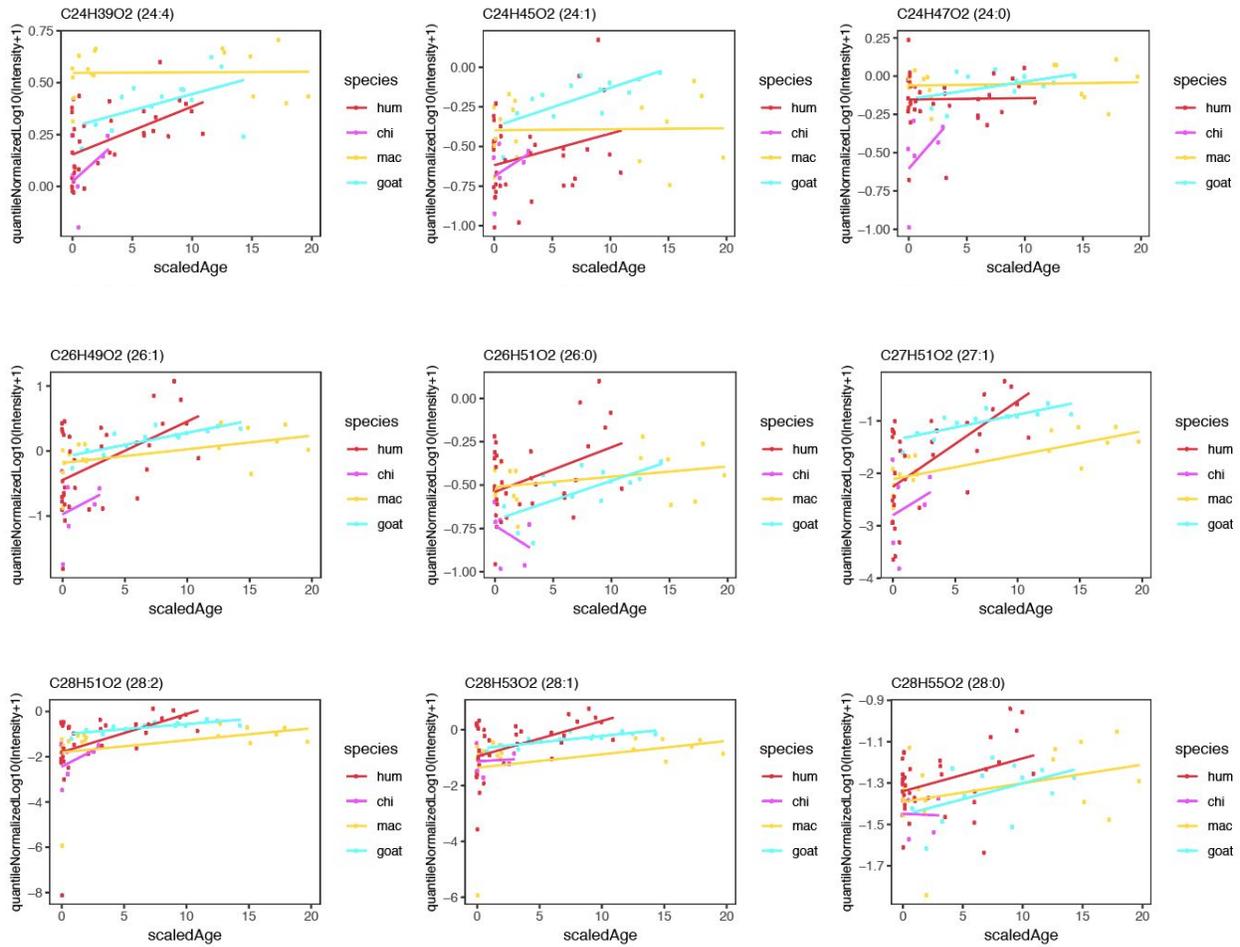


Fig. S3. Dependency of FA intensities in CB from the age. Color represents species. The x-axis represents scaled age in days. The y-axis represents upper-quartile normalized log-transformed signal intensity. Formula of the FA is specified on top.