Cancer Letters 452 (2019) 71–78

Contents lists available at ScienceDirect

Cancer Letters

journal homepage: www.elsevier.com/locate/canlet

Original Articles

Metabolomics of neonatal blood spots reveal distinct phenotypes of pediatric acute lymphoblastic leukemia and potential effects of early-life nutrition

Lauren M. Petrick[a,e], Courtney Schiffman[b,e], William M.B. Edmands[c], Yukiko Yano[c,e], Kelsi Perttula[c], Todd Whitehead[d,e], Catherine Metayer[d,e], Craig E. Wheelock[f], Manish Arora[a], Hasmik Grigoryan[c], Henrik Carlsson[c], Sandrine Dudoit[b,g], Stephen M. Rappaport[c,e,*]

[a] The Senator Frank R. Lautenberg Environmental Health Science Laboratory, Department of Environmental Medicine and Public Health, Icahn School of Medicine at Mount Sinai, New York, NY, USA
[b] Division of Biostatistics, School of Public Health, University of California, Berkeley, CA, USA
[c] Division of Environmental Health Sciences, School of Public Health, University of California, Berkeley, CA, USA
[d] Division of Epidemiology, School of Public Health, University of California, Berkeley, CA, USA
[e] Center for Integrative Research on Childhood Leukemia and the Environment, University of California, Berkeley, CA, USA
[f] Division of Physiological Chemistry 2, Department of Medical Biochemistry and Biophysics, Karolinska Institute, Stockholm, Sweden
[g] Department of Statistics, University of California, Berkeley, CA, USA

ARTICLE INFO

Keywords:
Lipids
Breastfeeding
Maternal BMI
Pre-B ALL
t(12;21) translocation

ABSTRACT

Early-life exposures are believed to influence the incidence of pediatric acute lymphoblastic leukemia (ALL). Archived neonatal blood spots (NBS), collected within the first days of life, offer a means to investigate small molecules that reflect early-life exposures. Using untargeted metabolomics, we compared abundances of small-molecule features in extracts of NBS punches from 332 children that later developed ALL and 324 healthy controls. Subjects were stratified by early (1–5 y) and late (6–14 y) diagnosis. Mutually-exclusive sets of metabolic features - representing putative lipids and fatty acids - were associated with ALL, including 9 and 19 metabolites in the early- and late-diagnosis groups, respectively. In the late-diagnosis group, a prominent cluster of features with apparent 18:2 fatty-acid chains suggested that newborn exposure to the essential nutrient, linoleic acid, increased ALL risk. Interestingly, abundances of these putative 18:2 lipids were greater in infants who were fed formula rather than breast milk (colostrum) and increased with the mother's pre-pregnancy body mass index. These results suggest possible etiologic roles of newborn nutrition in late-diagnosis ALL.

1. Introduction

Acute lymphoblastic leukemia (ALL) is the most common form of childhood cancer in the U.S. and other developed countries [1,2] with national incidence rates between < 20 and 60 cases per million children per year [3]. In a recent review of the etiology of childhood ALL, Greaves [4] summarized compelling evidence that the most common subtype of this disease, B cell precursor ALL (BCP-ALL), is caused by two distinct events. First, a pre-leukemic clone is initiated *in utero* by fusion-gene formation in approximately 1% of newborns. Then, about 1% of those children with pre-leukemic clones progress to overt leukemia, with a peak of incidence occurring at two to five years of age. Greaves concluded that the latency period (extending to ~15 years of age) for transition to BCP-ALL pointed to secondary genetic changes, notably those caused by early-life exposures to commensal microbes, infections and diet. Therefore, objective measures of early-life exposures and their biological imprints could point the way to discovering causes of ALL, at least in children diagnosed after one year of age [4,5].

Because microbiota, infections, diet and other potential environmental risk factors generate molecules that circulate in blood, a promising avenue for discovering causes of ALL involves comparisons of metabolomes between ALL cases and controls in prediagnostic blood

---

Abbreviations: ALL, acute lymphoblastic leukemia; BCP-ALL, B-cell precursor ALL; Hb, hemoglobin; FDR, false discovery rate; LC-HRMS, liquid chromatography high-resolution mass spectrometry; NBS, neonatal dried blood spots; CCLS, California Childhood Leukemia Study; lasso, least absolute shrinkage and selection operator; LysoPC, lysophosphatidylcholine; LysoPE, lysophosphatidylethanolamine; MS2, tandem MS/MS fragmentation; PC, phosphadityIcholine; PS, phosphatidylserine; PUFA, polyunsaturated fatty acid; SM, sphingomyelin; t(12, 21), translocation also known as ETV6–RUNX1 and TEL-AML1

[*] Corresponding author. Center for Integrative Research on Childhood Leukemia and the Environment, University of California, 1995 University Ave., Suite 460, Berkeley, CA, 94704, USA.

E-mail address: srappaport@berkeley.edu (S.M. Rappaport).

https://doi.org/10.1016/j.canlet.2019.03.007
Received 28 December 2018; Received in revised form 1 March 2019; Accepted 10 March 2019

[6]. Archived neonatal blood spots (NBS, also known as Guthrie cards), which are collected from virtually all live births in the U.S. to test for metabolic defects, offer avenues for detecting exposures that occur near birth. Since 1982 the State of California has archived unused NBS at −20 °C for epidemiological investigations [7].

We recently developed an untargeted metabolomics method with NBS via liquid chromatography high-resolution mass spectrometry (LC-HRMS) [8]. Here we report results from analyses of 656 archived NBS from ALL cases, diagnosed after one year of age, and matched controls who participated in the California Childhood Leukemia Study (CCLS) [9]. Because the age at diagnosis has been shown to affect the strength of associations with several ALL risk factors [10–12], we stratified case-control pairs by early diagnosis (1–5 years) coinciding with peak incidence, and late diagnosis (6–14 years) [4,13]. Data were filtered and normalized, focusing on 869 abundant features that were detected in most NBS.

2. Materials and methods

2.1. Neonatal blood spots

The CCLS is a case–control study conducted between 1995 and 2015 to identify risk factors for childhood leukemia. Incident cases of childhood leukemia, diagnosed up to 14 years of age, were ascertained across California, generally within 72 h of diagnosis [9]. Archived NBS for CCLS participants were obtained from the California biobank program (Sacramento, CA). We used 4.7-mm punches (equivalent to ∼8 μL of whole blood) from 656 NBS collected between 1985 and 2005, with ALL cases and controls matched by date of birth, sex, ethnicity (one or both parents being Hispanic) and maternal race. Information on sociodemographic characteristics was obtained from parent interviews. Summary statistics are available in Supplementary Table 1, available as Supplementary Materials. An additional set of blank punches was obtained from adjacent portions of the same Guthrie cards.

2.2. Metabolomic analysis

A total of 656 NBS punches were analyzed (Supplementary Table 1). Briefly, samples were extracted with water and assayed for potassium [8] (batch 1) or hemoglobin (Hb) [14] (batches 2–4) to adjust for blood volume (see Supplementary Fig. 1). Then, acetonitrile was added to precipitate proteins and extracts were analyzed by LC-HRMS [8]. Data processing was performed in the R statistical programming environment using methods described elsewhere [15]. Detection of sample outliers, beyond a proportional expansion value of 1.2 for Hotelling's ellipse (PC1 and PC2), resulted in removal of two cases and 10 controls. Of the 61,945 features detected in NBS and blank punches, filtering features by blank samples left 25,261, excluding features with more than 20% missing values left 1,606, and excluding features with intraclass correlation coefficients less than 0.2 left 869 features to be examined for their associations with ALL. Missing values were imputed abundances based on k-nearest neighbor imputation using $k = 5$. Features were annotated by comparing masses, isotope patterns, and MSMS fragmentation spectra [16,17], and confirmed (when possible) with authentic standards. Annotation confidence was evaluated with criteria reported by Schymanski et al. [18] (Table 1).

2.3. Feature selection

Feature selection was performed separately for early- and late-diagnosis of ALL. Peak areas were log transformed and normalized with the Bioconductor R package ‘scone’ [15,19,20], which implemented and evaluated different scaling and regression-based-normalization methods for removing unwanted variation while preserving differences in case status. The normalization scheme selected by ‘scone’ used DESeq scaling and accounted for the following unwanted sources of variation: *NBS age*, *blood volume*, *run order*, and *batch*.

To capture different types of associations between metabolites and ALL, features were selected using a combination of methods based on multivariate linear regression, regularized logistic regression and random forest [21]. First, the following linear regression model was used for a given logged and scaled feature Y in the ith subject:

$$Y_i = \beta_0 + \beta_1 X_{i,caco} + \beta_2 X_{i,NBS\ age} + \beta_3 X_{i,sex} + \beta_4 X_{i,ethnicity} + \beta_5 X_{i,blood\ volume} + \beta_6 X_{i,run\ order} + \beta_7 X_{i,batch} + \varepsilon_i, \quad [1]$$

where: *caco* denotes the binary case-control status, *NBS age*, *sex*, and *ethnicity* are matching variables, *blood volume* represents the volume of blood in the NBS punch, and *run order* and *batch* adjust for technical variation. Features were ranked by their nominal unadjusted p-values for the case-control coefficient (β_1), and the case/control fold change was estimated as exp(β_1). Second, a regularized logistic regression (lasso) [22,23] model was fit to normalized feature abundances over 500 bootstrapped datasets, with *caco* as the outcome variable and the following independent variables: normalized logged intensities for all 869 features and matching variables (*NBS age*, *sex* and *ethnicity*). The percentage of time that each feature was selected by lasso across the bootstrap iterations was used to rank the association with ALL. Features that were ranked in the top 5% for both linear regression p-value and lasso were joined with those of high importance from random forest (separated by more than a 20% increase in importance) to include possible non-linear associations with ALL [24–26]. After removing selected features with poor peak morphology, ion suppression, as well as minor isotopes and electrospray adducts, the ensemble of variable selection methods was repeated iteratively until a final set of 28 features was selected, 9 for early diagnosis and 19 for late diagnosis. Correlations between normalized abundances of the selected features were displayed with agglomerative hierarchical clustering using complete linkage and Spearman correlation (‘hclust’ function in R). Apparent clusters from hierarchical clustering were validated with the partitioning around medoids (PAM) method (k = 2, …, 6 ‘pam’ function in R) also using Spearman correlations. Associations between selected features and covariates were visualized using scatter plots with loess smoothing.

2.4. Evaluation of covariates

To investigate factors that could potentially confound relationships between predictive features and ALL status, the continuous covariates *breastfeeding duration* (weeks), *birthweight* (g) and *mother's BMI* (kg/m^2; prenatal) were evaluated because breastfeeding, maternal diet, gestational diabetes, and birthweight have been suggested as risk factors for ALL [27,28]. Confounding by household income (*income,* binary, with a cut point of \$60,000) was also considered because the number of cases with low income ($n = 185$) was higher than that of controls ($n = 126$) (Supplementary Table 1). To investigate possible confounders and relative strengths of associations with ALL, a random forest model was used to rank all selected features and covariates by their importance in predicting ALL [29,30].

A case-case analysis of features predictive of t(12; 21) was investigated in the early diagnosis group where 44 of the 214 cases (21%) had this cytogenetic abnormality (Supplementary Table 1). Additional stratification of the early or late diagnosis group was performed to investigate potential confounding by the pre-B cell phenotype [cases with t(12; 21) translocations, hyperdiploidy and T-cell phenotypes].

3. Results

3.1. Features associated with ALL

Table 1 lists all features, identified by accurate mass (m/z value), that were selected for associations with either early diagnosis ($n = 9$) or

Table 1
Annotation of metabolites associated with early diagnosis and late diagnosis of ALL.

Accurate mass (*m/z*)[a]	Name[b]	CL[c]	Chemical class	Fold change[d]	*p*-value[e]	Cluster[f]
Early diagnosis (1–5 years)						
363.3263	Tetracosadienoic acid	4	Unsaturated fatty acid	0.94	0.0027	1
431.3151	Unknown	5		1.12	0.0063	2
570.3401	Unknown	5		1.09	0.0072	2
604.3610	LysoPC(20:3)	4	Glycerophospholipid	1.07	0.0489	2
638.2760	Unknown	5		1.08	0.0034	2
696.3573	Unknown	5		1.07	0.0177	2
766.5589	PS(16:0/16:0)	4	Glycerophospholipid	1.08	0.0037	2
775.2985	Unknown	5		1.09	0.0164	2
884.6004	PC(18:0/20:4 (OH))	3	Oxidized glycerophospholipid	1.09	0.0210	2
Late diagnosis (6–14 years)						
251.2011	Hexadecadienoic acid	4	Unsaturated fatty acid	0.89	0.0097	5
277.2170	Linolenic acid[g]	1	Unsaturated fatty acid	1.22	0.0016	3
279.2329	Linoleic acid	1	Unsaturated fatty acid	1.14	0.0092	3
377.1419	Unknown	5		1.20	0.0102	3
387.3264	$C_{26}H_{44}O2$	4		1.07	0.0548	4
449.3629	Unknown	5		0.91	0.0530	5
476.2778	LysoPE(18:2)	2	Glycerophospholipid	1.18	0.0032	3
500.2782	LysoPE(20:4)	2	Glycerophospholipid	1.07	0.0222	4
530.3234	Unknown PE	5	Glycerophospholipid	1.14	0.0161	4
552.2956	Unknown	5		0.91	0.0441	5
564.5344	$C_{36}H_{71}NO_3$	4	Sphingolipid	0.90	0.0390	5
578.3458	LysoPC (18:2)	2	Glycerophospholipid	1.22	0.0064	3
592.3238	Unknown	5		0.91	0.0391	5
760.5126	PS(16:0/18:1)	3	Glycerophospholipid	1.08	0.0387	4
789.6123	SM(d16:1/20:0)	3	Sphingolipid	0.95	0.0497	5
824.5793	PC(P16:0/20:4)	4	Glycerophospholipid	1.07	0.0519	4
844.6069	PC(18:0/18:2)	3	Glycerophospholipid	1.10	0.0154	3
897.6298	Unknown	5		0.93	0.0381	5
965.7626	$C_{64}H_{102}O_6$	4		1.16	0.0464	3

[a] Feature identifier given by accurate mass (*m/z*).
[b] Common names as used in the Human Metabolome Database (HMDB): lysophosphatidylcholine, LysoPC; phosphaditylcholine, PC; lysophosphatidylethanolamine, LysoPE; phosphatidylserine, PS; sphingomyelin, SM.
[c] Confidence level of annotation consistent with the scheme of Schymanski et al. [18] (1 = highest, 5 = lowest confidence).
[d] Case/control fold change of feature abundances.
[e] Nominal *p*-value for the case-control coefficient (β_1) from Model 1.
[f] Correlation cluster (Fig. 1).
[g] It was not possible to distinguish between alpha- and gamma-isomers.

late diagnosis ($n = 19$) of ALL. Effect sizes were modest, with case/control fold changes ranging from 0.94 to 1.11 for early diagnosis and from 0.89 to 1.22 for late diagnosis. Based on annotations (Table 1 and Supplementary Table 2), features that discriminated ALL cases from controls were lipids and unsaturated fatty acids.

3.2. Correlations of features

Heat maps were used to visualize clusters of features associated with ALL in the early- and late-diagnosis groups (Fig. 1A and B, respectively). Results for PAM clustering were consistent (Supplementary Fig. 2). For the early-diagnosis group two clusters were observed (C1 and C2 in Fig. 1 and Table 1). Cluster C1 contained 8 features, all of which were more abundant in cases than controls (Table 1); three of these features were putatively annotated as glycerophospholipids (604.3610, 766.5589, and 884.6004). Only feature 363.3263 (putative tetracosadienoic acid) in cluster C2 was present at lower levels in cases.

There were three clusters in the late-diagnosis group (C3 – C5 in Fig. 1B and Table 1). All features in clusters C3 and C4 were more abundant in cases while all of those in C5 were less abundant in cases. Annotations in clusters C3 and C4 included linolenic acid (277.2170), linoleic acid (279.2329) and several putative glycerophospholipids (476.2778, 500.2782, 530.3234, 578.3458, 760.5126, 824.5793 and 844.6069). Many putative metabolites in cluster C3 contained 18:2 fatty acid chains, including unequivocally identified linoleic acid. Thus, we speculate that unknown metabolites 377.1419 and 965.7627 may also contain 18:2 moieties. Cluster C4 contained two putative metabolites with arachidonic acid side chains (20:4). The putative annotations in cluster C5 included sphingolipids (789.6123 and 564.5344), and a metabolite of conjugated linoleic acid (hexadecadienoic acid, 251.2011).

3.3. Correlations with covariates

Random forest variable-importance plots, shown in Fig. 2, indicate that *mother's BMI, birthweight* and *breastfeeding duration* ranked more highly for classifying ALL cases than *income* for both early- and late-diagnosis, but were ranked lower than all metabolomic features. This suggests that predictive metabolites were more discriminating for case status than any of the tested covariates and that selection of the predictive metabolites was not unduly influenced by these covariates. Nonetheless, the heat maps in Fig. 1 show that *breastfeeding duration* and *mother's BMI* were consistently correlated with six features in cluster C3, including those with known or putative 18:2 fatty acid moieties. Interestingly, the directions of these correlations were reversed for all six features, i.e., negative correlations with *breastfeeding duration* and positive correlations with *mother's BMI.* Scatter plots with loess smoothing further elucidated these relationships, as shown in Fig. 3.

3.4. Stratified analyses

ALL cases at early-diagnosis had a higher percentage of t(12; 21) translocations (46/205 = 21%) than those at late-diagnosis (6/106 = 5%) (Supplementary Table 1). To assess the associations between t(12; 21) and the 9 features predictive of early diagnosis, the

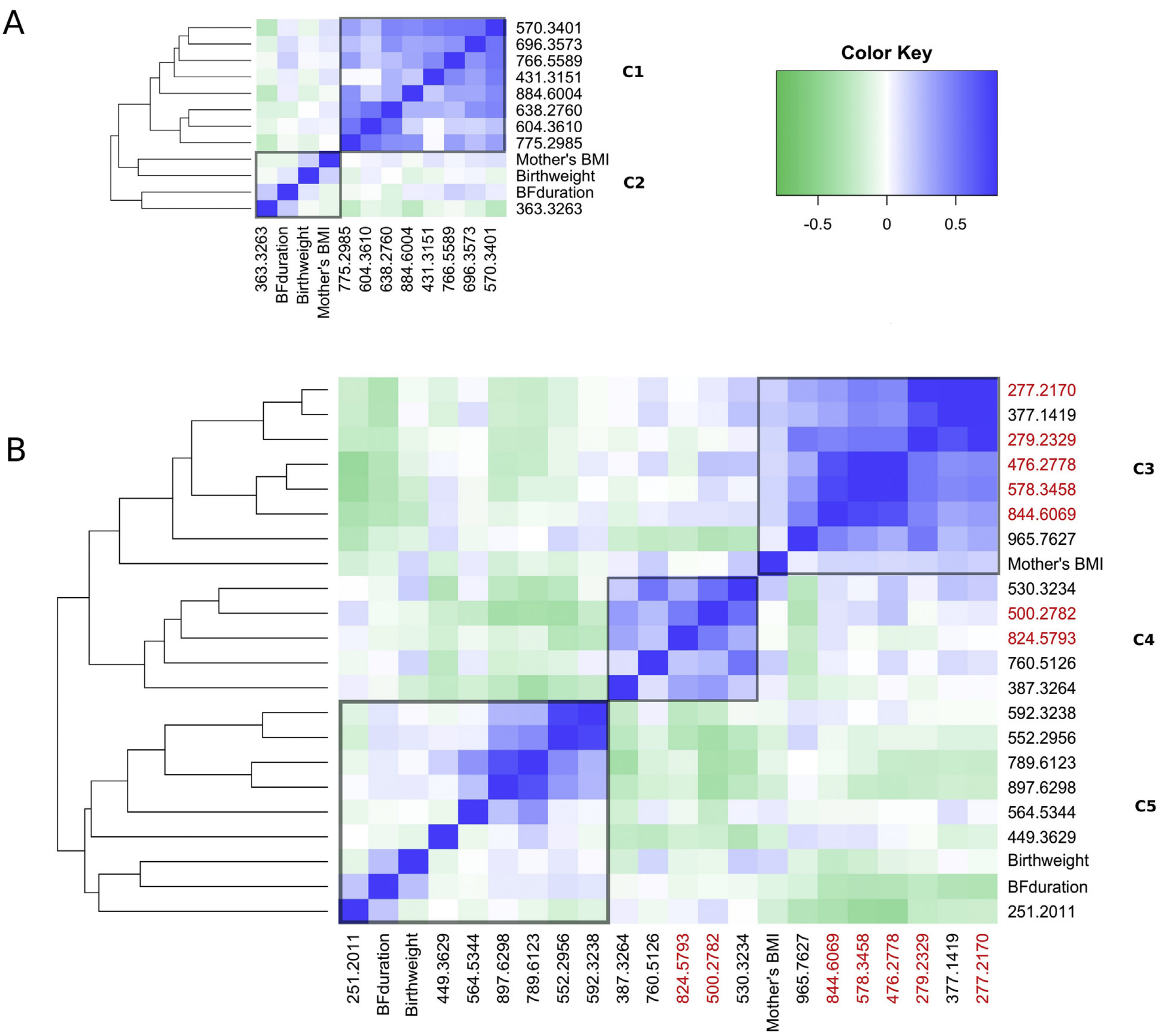

Fig. 1. Agglomerative hierarchical clustering using complete linkage and Spearman correlation ('hclust' function in R). Clusters of features predictive of (A) early diagnosis and (B) late diagnosis of ALL, with distinct clusters labeled C1-C5. Metabolites containing 18:2, 18:3 or 20:4 fatty acid chains are highlighted in red.

variable selection method was repeated by comparing cases with and without t(12; 21). This resulted in 13 features predictive of t(12; 21) (Supplementary Table 3). Only feature 604.3610 [LysoPC(20:3)] overlapped with those in Table 1 for the early-diagnosis group.

Removing cases with T-cell leukemia (19 in total, 8 from the early group and 11 from the late group) from the analysis did not change the results.

Removing cases with 'early pre-B' cell phenotypes [those with t(12; 21) translocations, hyperdiploidy, and T-cell phenotypes] from the late-diagnosis group resulted in 63 cases with 'other B-cell' phenotypes (Supplementary Table 1). Comparing these to controls ($n = 117$) slightly improved effect sizes as determined by Model 1 for 17 out of 19 metabolites predictive of late diagnosis (0.1–6.3%). This increase was not observed when evaluated in the early-diagnosis group.

4. Discussion

Because ALL risks had previously been shown to be affected by age at diagnosis [10–13], we stratified cases by early (1–5 years) and late diagnosis (6–14 years) and, indeed, discovered mutually-exclusive sets of predictive metabolomic features (Table 1). These 28 metabolites were mainly putative lipids (Table 1), some of which have been found to be perturbed in diagnostic blood for a number of malignancies [31,32] including childhood and adult acute leukemias [33,34]. Our study is unique in that the findings are based on pre-diagnostic blood collected at birth.

Almost all of the putative glycerophospholipids predictive of ALL, including PCs and LysoPCs, were more abundant in cases than controls (Table 1). This finding is consistent with reports that PCs and LysoPCs are overexpressed in some cancers [35,36] and can influence both cell proliferation and apoptosis [37,38]. Also, choline modulation has been shown to be a source of PCs and LysoPCs in diagnostic blood from AML cases compared to controls and other tumors [39]. At the cellular level, LysoPCs have been reported to be more abundant in CCRF-CEM leukemia cells after drug treatment due to increased oxidative stress [40]. Interestingly, a putative oxidized PC [PC(18:0/20:4(OH), 884.6004] that was more abundant in early ALL cases (Table 1) represents a class of biomarkers of oxidative stress [41,42].

Nine metabolites were predictors of early-ALL diagnosis and all but one were more abundant in cases than controls (Table 1). The eight positively-associated features were correlated (cluster C1 of Fig. 1A), and those with putative annotations were products of glycerophospholipid metabolism. The other feature, 363.3263 (putative tetracosadienoic acid, cluster C2 of Fig. 1A), was less abundant in cases than

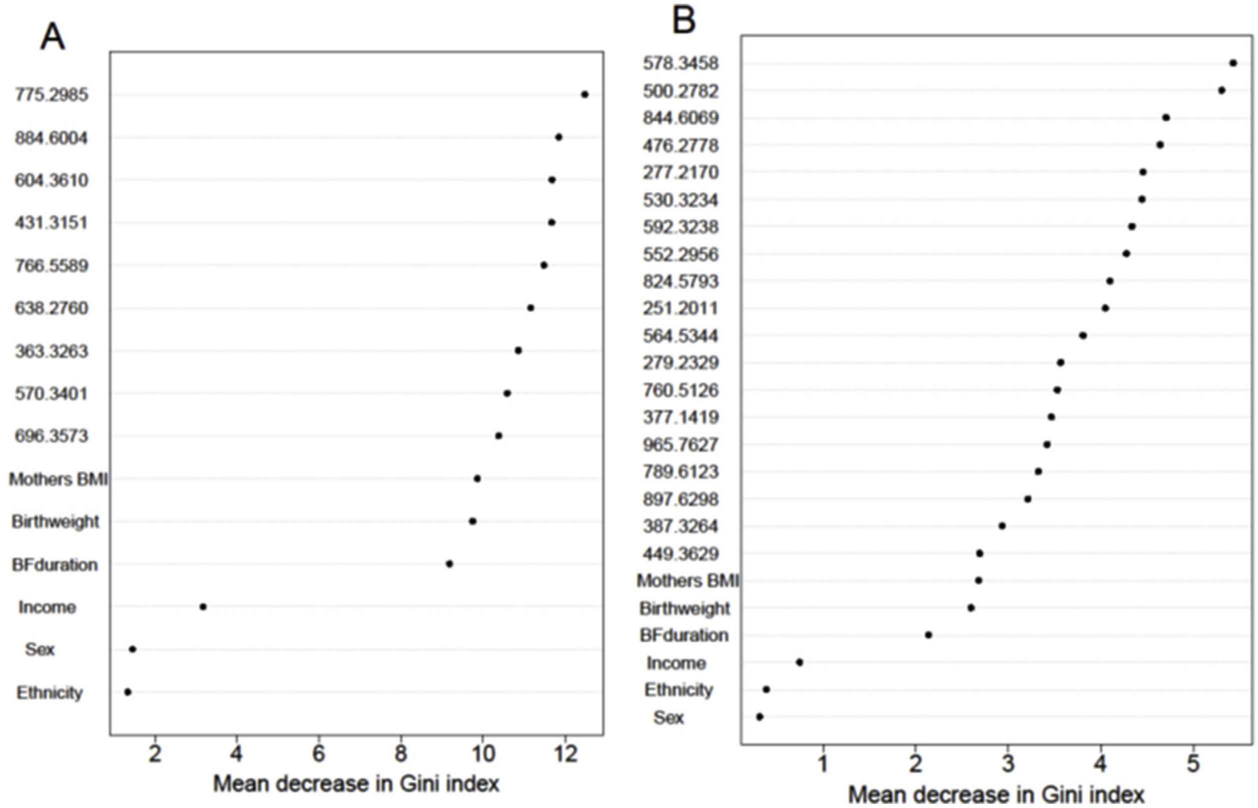

Fig. 2. Random Forest variable importance plots for (A) early diagnosis and (B) late diagnosis of ALL. Legend: BF duration, breastfeeding duration.

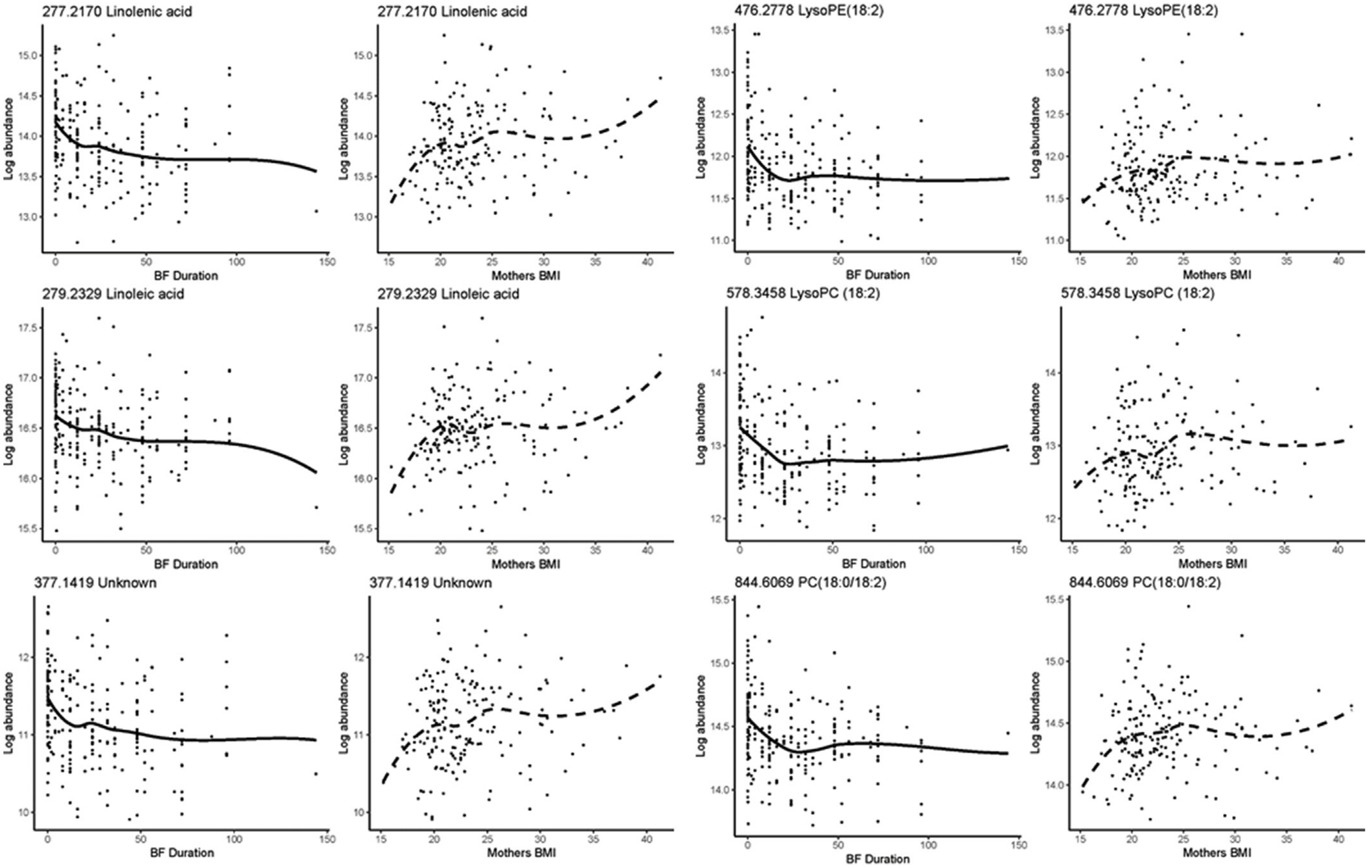

Fig. 3. Scatter plots with loess smoothing of feature abundances from cluster C3 that were correlated with breastfeeding (BF) duration (weeks) and mother's pre-pregnancy body mass index (BMI, kg/m^2).

controls and was positively correlated with *breastfeeding duration* (Table S3).

In the late-diagnosis group, 19 metabolomic features discriminated ALL cases from controls, including a putative ceramide (564.5344) and sphingomyelin [SM(d16:1/20:0), 789.6123)], which were less abundant in cases (Table 1) and were correlated with each other (C5, Fig. 1B). This could reflect the effects of acid ceramidase, which catalyzes ceramide breakdown and has been shown to be overexpressed in AML [43], and of altered sphingolipid metabolism that has been implicated with cancer progression [44].

We find it interesting that two essential fatty acids, linoleic acid (18:2n6, 279.2329) and linolenic acid (277.2170), were more abundant in late-diagnosis cases (Table 1), suggesting that maternal and/or newborn nutrition were involved in the early-life etiology of ALL. This conjecture is supported by correlations of putative features in cluster C3, including several with 18:2 fatty acid chains, with *mother's BMI* and with *breastfeeding duration*. As shown by the scatter plots in Fig. 3, abundances of the features in cluster C3 increased with *mother's BMI*, reinforcing a previous finding that newborns' levels of linoleic acid were positively correlated with the mothers' BMI [45]. The negative correlations between *breastfeeding duration* and the same putative 18:2 features in cluster C3 (Fig. 1), are informative because NBS were typically obtained between 24 and 48 h post-delivery after infants had received multiple feedings of either breastmilk (colostrum) or formula. Since levels of linoleic and linolenic acid have been shown to be lower in colostrum than formula [46–48], post-delivery breast feeding arguably led to reduced abundances of these 18:2 and 18:3 fatty acids in NBS and also served as sentinels of *breastfeeding duration*, which has been shown to reduce risks of ALL in the CCLS and other studies [12,28,49,50]. Conversely, *breastfeeding duration* was positively correlated with putative hexadecadienoic acid (16:2n6, 251.2011), a metabolite of conjugated linoleic acid that has been shown to be anti-carcinogenic [51,52] and is more abundant in colostrum than formula [46,53]. The gut microbiome, including *Bifidobacterium* and *Lactobacillus*, is involved in the conversion of linoleic acid to conjugated linoleic acid [54].

It is worth mentioning that Shu et al. [55] observed a lower odds ratio for the association of breastfeeding with ‘pre-B ALL’ than ‘early pre-B ALL’ phenotypes. Since ‘early pre-B ALL’ includes B cell cases with t(12; 21) translocations and hyperdiploidy, we may be observing a similar effect because removal of t(12; 21), hyperdiploidy, and T-cell ALL cases from the late diagnosis group led to modest increases in the effect sizes of the same metabolites. Scatter plots (Fig. 3) also show that the relationships between feature abundances and *breastfeeding duration* occurred predominately in the first 26 weeks (< 6 months). Thus, the correlations between metabolites selected for ALL in the late-diagnosis group and *breastfeeding duration* may be driven by a specific phenotype consistent with the findings of Shu et al.

Because most of the predictive metabolites of late-diagnosis ALL were positively correlated (Fig. 1B), they may represent a single underlying pathway or network. The predominant pathways associated with late diagnosis had case/control fold changes greater than one (Table 1) and were related to parent (linoleic and linolenic) fatty acids, while the corresponding conjugated linoleic acid metabolite (putative hexadecadienoic acid) had a fold change less than one. Higher fold changes were also observed for structural lipids, including putative PC, LysoPC and LysoPE species that contained linoleic (18:2n6) and arachidonic acid (20:4n6) side chains. Fatty acids, including linoleic acid (n-6) are converted by desaturases to long-chain PUFAs including arachidonic acid [56]. Increased arachidonic acid levels can result in increased eicosanoid production (e.g., prostaglandins), which can influence cancer progression, possibly through an immune response [57,58]. These findings suggest that early nutritional intake in the form of fatty acid consumption is associated with ALL and may involve the downstream biosynthetic machinery including desaturase/elongase enzymes and/or eicosanoid synthesis.

We recognize that this discovery study is limited to a single sample of ALL cases and matched controls and will require validation in independent cohorts. Annotations of the lipid features predictive of ALL in our study were limited by the LC-HRMS platform used for untargeted metabolomics, and by online databases that are particularly lacking in negative mode data, especially with acetic acid as an additive. Because of this, only linoleic acid and linolenic acid were unambiguously identified based on comparisons with reference standards. We encourage future studies to employ methods that can discriminate fatty acid isomers (e.g., gas chromatography–mass spectrometry).

In summary, fetal metabolomics of NBS revealed putative lipid modifications associated with childhood leukemia that differed between early and late diagnosis of ALL, notably lipids containing 18:2 moieties derived from dietary linoleic acid that were more abundant in late-diagnosis cases than controls. Interestingly, these same putative lipids were negatively correlated with breastfeeding duration, thus supporting epidemiological findings that breastfeeding is protective for ALL. This work should encourage efforts to elucidate systems biology that links lipidomic pathways with early-life nutrition and the associated ALL risks.

Conflicts of interest

The authors declare no conflict of interest.

Ethics approval

The study was approved by the University of California Committee for the Protection of Human Subjects, the California Health and Human Services Agency Committee for the Protection of Human Subjects, and the institutional review boards of all participating hospitals.

Informed consent

Written informed consent was obtained from the parents of all participating subjects.

CRediT author statement

Lauren M. Petrick: Methodology, Investigation, Writing-Original Draft, Writing-Review & Editing; **Courtney Schiffman**: Formal Analysis, Software, Data Curation, Visualization, Writing-Review & Editing; **William M.B. Edmands**: Software; **Yukiko Yano**: Methodology, Investigation, Writing-Review & Editing; **Kelsi Perttula**: Investigation; **Todd Whitehead**: Resources, Data Curation, Writing-Review & Editing, **Catherine Metayer**: Supervision, Funding Acquisition, Project Administration, Writing-Review & Editing; **Craig E. Wheelock**: Formal Analysis, Writing-Review & Editing, **Manish Arora**: Resources, Supervision; **Hasmik Grigoryan**: Methodology, **Henrik Carlsson**: Investigation, **Sandrine Dudoit**: Methodology, Software, Formal Analysis, Writing-Review & Editing, **Stephen M. Rappaport**: Conceptualization, Formal Analysis, Writing-Original Draft, Writing-Review & Editing, Supervision, Project Administration, Funding Acquisition.

Funding

This work was supported by the National Institute for Environmental Health Sciences at the National Institutes of Health [P01 ES018172, P50 ES018172, P01 ES018172, U2C ES026561]; the United States Environmental Protection Agency [RD83615901, RD83451101]; Children with Cancer U.K. [partial support for cytogenetic characterization], as well as a post-doctoral fellowship by the Environment and Health Fund, Israel. C.E.W. was supported by the Swedish Heart-Lung Foundation [20170736, 20170603]. The biospecimens used in this study were obtained from the California Biobank Program, California

Department of Public Health (SIS request number 26), in accordance with Section 6555(b), 17 CCR. The NIEHS, USEPA, and California Department of Public Health are not responsible for the results or conclusions drawn by the authors of this publication.

Acknowledgements

We appreciate the generous loan of LC-HRMS equipment by Agilent Technologies (Santa Clara, CA). We thank the families for their participation. We also thank the clinical investigators at the following collaborating hospitals for help in recruiting patients: University of California Davis Medical Center (Dr. Jonathan Ducore), University of California San Francisco (Drs. Mignon Loh and Katherine Matthay), Children's Hospital of Central California (Dr. Vonda Crouse), Lucile Packard Children's Hospital (Dr. Gary Dahl), Children's Hospital Oakland (Dr. James Feusner), Kaiser Permanente Roseville (former Sacramento) (Drs. Kent Jolly and Vincent Kiley), Kaiser Permanente Santa Clara (Drs. Carolyn Russo, Alan Wong and Denah Taggar), Kaiser Permanente San Francisco (Dr. Kenneth Leung) and Kaiser Permanente Oakland (Drs. Daniel Kronish and Stacy Month). We also thank the entire California Childhood Leukemia Study staff.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.canlet.2019.03.007.

References

[1] Cancer Facts & Figures 2014: Special Addition, (n.d.). http://www.cancer.org/acs/groups/content/@research/documents/webcontent/acspc-041787.pdf (accessed March 4, 2016).

[2] H. Inaba, M. Greaves, C.G. Mullighan, Acute lymphoblastic leukaemia, Lancet 381 (2013) 1943–1955, https://doi.org/10.1016/S0140-6736(12)62187-4.

[3] W.H.O.Europe, Incidence of childhood leukaemia, Eur. Environ. Heal. Inf. Syst. 2009 (2000) 5 doi:Code: RPG4_Rad_E1.

[4] M. Greaves, A causal mechanism for childhood acute lymphoblastic leukaemia, Nat. Rev. Canc. (2018) 1–14, https://doi.org/10.1038/s41568-018-0015-6.

[5] P. Brown, Treatment of infant leukemias: challenge and promise, Hematology Am. Soc. Hematol. Educ. Program 2013 (2013) 596–600, https://doi.org/10.1182/asheducation-2013.1.596.

[6] S. Rappaport, Redefining environmental exposure for disease etiology, NPJ Syst. Biol. Appl. 4 (2018).

[7] California Department of Public Health, Background and History of the California Biobank Program (CBP), (2016) https://www.cdph.ca.gov/programs/GDSP/Pages/MoreAboutTheCBP.aspx.

[8] L. Petrick, W. Edmands, C. Schiffman, H. Grigoryan, K. Perttula, Y. Yano, S. Dudoit, T. Whitehead, C. Metayer, S. Rappaport, An untargeted metabolomics method for archived newborn dried blood spots in epidemiologic studies, Metabolomics 13 (2017) 1–11, https://doi.org/10.1007/s11306-016-1153-z.

[9] C. Metayer, L. Zhang, J.L. Wiemels, K. Bartley, J. Schiffman, X. Ma, M.C. Aldrich, J.S. Chang, S. Selvin, C.H. Fu, J. Ducore, M.T. Smith, P. a Buffler, Tobacco smoke exposure and the risk of childhood acute lymphoblastic and myeloid leukemias by cytogenetic subtype, Cancer Epidemiol. Biomark. Prev. 22 (2013) 1600–1611, https://doi.org/10.1158/1055-9965.EPI-13-0350.

[10] O. Paltiel, G. Tikellis, M. Linet, J. Golding, S. Lemeshow, G. Phillips, K. Lamb, C. Stoltenberg, S.E. Håberg, M. Strøm, C. Granstrøm, K. Northstone, M. Klebanoff, A.L. Ponsonby, E. Milne, M. Pedersen, M. Kogevinas, E. Ha, T. Dwyer, Birthweight and childhood cancer: preliminary findings from the international childhood cancer cohort consortium (I4C), Paediatr. Perinat. Epidemiol. 29 (2015) 335–345, https://doi.org/10.1111/ppe.12193.

[11] A. Wallace, W. Francis, X. Ma, R. McKean-Cowdin, S. Selvin, T. Whitehead, L. Barcellos, A. Kang, L. Morimoto, T. Moore, J. Wiemels, C. Metayer, Allergies and Childhood Acute Lymphoblastic Leukemia: A Case-Control Study and Meta-Analysis, CEBP, 2018.

[12] J. Rudant, T. Lightfoot, K.Y. Urayama, E. Petridou, J.D. Dockerty, C. Magnani, E. Milne, L.G. Spector, L.J. Ashton, N. Dessypris, A.Y. Kang, M. Miller, R. Rondelli, J. Simpson, E. Stiakaki, L. Orsi, E. Roman, C. Metayer, C. Infante-Rivard, J. Clavel, Childhood acute lymphoblastic leukemia and indicators of early immune stimulation: a childhood leukemia international consortium study, Am. J. Epidemiol. 181 (2015) 549–562, https://doi.org/10.1093/aje/kwu298.

[13] T. Westergaard, P.K. Andersen, J.B. Pedersen, J.H. Olsen, M. Frisch, H.T. Sorensen, J. Wohlfahrt, M. Melbye, Birth characteristics, sibling patterns, and acute leukemia risk in childhood: a population-based cohort study, J. Natl. Cancer Inst. 89 (1997) 939–947, https://doi.org/10.1093/jnci/89.13.939.

[14] Y. Yano, H. Grigoryan, C. Schiffman, W.M. Edmands, L. Petrick, K. Hall, T. Whitehead, C. Metayer, S. Rappaport, Untargeted adductomics of Cys34 modifications to human serum albumin in newborn dried blood spots, Anal. Bioanal. Chem. (2019 Feb 19), https://doi.org/10.1007/s00216-019-01675-8 [Epub ahead of print] PMID:30783713.

[15] C. Schiffman, L. Petrick, K. Perttula, Y. Yano, H. Carlsson, T. Whitehead, C. Metayer, J. Hayes, W.M.B. Edmands, S. Rappaport, S. Dudoit, Data-adaptive pipeline for filtering and normalizing metabolomics data, BioRxiv (2018), p. 387365, https://doi.org/10.1101/387365.

[16] B. Brugger, G. Erbent, R. Sandhoff, F.T. Wieland, W. Lehmann, Quantitative analysis of biological membrane lipids at the low picome level by nano-electrospray ionization tandem mass spectrometry, Proc. Natl. Acad. Sci. Unit. States Am. 94 (1997) 2339–2344.

[17] W.M.B. Edmands, L. Petrick, D.K. Barupal, A. Scalbert, M.J. Wilson, K. Wickli, S.M. Rappaport, compMS2Miner, An automatable metabolite identification, visualization, and data-sharing R package for high-resolution LC − MS data sets, Anal. Chem. 89 (2017) 3919–3928, https://doi.org/10.1021/acs.analchem.6b02394.

[18] E.L. Schymanski, J. Jeon, R. Gulde, K. Fenner, M. Ru, H.P. Singer, J. Hollender, Identifying small molecules via high resolution mass spectrometry: communicating confidence, Environ. Sci. Technol. 48 (2014) 2097–2098, https://doi.org/10.1021/es5002105 |.

[19] M.B. Cole, D. Risso, A. Wagner, D. DeTomaso, J. Ngai, E. Purdom, S. Dudoit, N. Yosef, Performance Assessment and Selection of Normalization Procedures for Single-Cell RNA-Seq, (2017), p. 235382, https://doi.org/10.1101/235382 BioRxiv.

[20] D. Risso, J. Ngai, T.P. Speed, S. Dudoit, Normalization of RNA-seq data using factor analysis of control genes or samples, Nat. Biotechnol. 32 (2014) 896–902, https://doi.org/10.1038/nbt.2931.

[21] K. Perttula, C. Schiffman, W.M. Edmands, L. Petrick, H. Grigoryan, X. Cai, M.J. Gunter, A. Naccarati, S. Polidoro, S. Dudoit, P. Vineis, S. Rappaport, Untargeted lipidomic features associated with colorectal cancer in a prospective cohort, BMC Canc. 18 (1) (2018 Oct 19) 996, https://doi.org/10.1186/s12885-018-4894-4.

[22] F. Bach, Bolasso: model consistent Lasso estimation through the bootstrap, ArXive (2008), https://doi.org/10.1145/1390156.1390161.

[23] J. Friedman, T. Hastie, R. Tibshirani, J. Stat. Softw. 33 (2010) 1–22 https://www.jstatsoft.org/index.

[24] H. Wang, F. Yang, Z. Luo, An experimental study of the intrinsic stability of random forest variable importance measures, BMC Bioinf. 17 (2016) 1, https://doi.org/10.1186/s12859-016-0900-5.

[25] M.L. Calle, V. Urrea, Letter to the editor: stability of random forest importance measures, Briefings Bioinf. 12 (2011) 86–89, https://doi.org/10.1093/bib/bbq011.

[26] L. Breiman, Random forests, Mach. Learn. 45 (2001) 5–32, https://doi.org/10.1023/A:1010933404324.

[27] N. Dessypris, M.A. Karalexi, E. Ntouvelis, A.A. Diamantaras, V. Papadakis, M. Baka, E. Hatzipantelis, M. Kourti, M. Moschovi, S. Polychronopoulou, V. Sidi, E. Stiakaki, E.T. Petridou, Association of maternal and index child's diet with subsequent leukemia risk: a systematic review and meta analysis, Cancer Epidemiol 47 (2017) 64–75, https://doi.org/10.1016/j.canep.2017.01.003.

[28] E.L. Amitay, G. Dubnov Raz, L. Keinan-Boker, Breastfeeding, Other early life exposures and childhood leukemia and lymphoma, Nutr. Canc. 68 (2016) 968–977, https://doi.org/10.1080/01635581.2016.1190020.

[29] S. Schneeweiss, W. Eddings, R.J. Glynn, E. Patorno, J. Rassen, J.M. Franklin, Variable selection for confounding adjustment in high-dimensional covariate spaces when analyzing healthcare databases, Epidemiology 28 (2017) 237–248, https://doi.org/10.1097/EDE.0000000000000581.

[30] M. Lu, S. Sadiq, D. Feaster, H. Ishwaran, Estimating individual treatment effect in observational data using random forest methods, J. Comput. Graph. Stat. 17 (2018) 209–219, https://doi.org/10.1016/j.chemosphere.2012.12.037.Reactivity.

[31] E.G. Armitage, A.D. Southam, Monitoring cancer prognosis, diagnosis and treatment efficacy using metabolomics and lipidomics, Metabolomics 12 (2016) 1–15, https://doi.org/10.1007/s11306-016-1093-7.

[32] R. Bandu, J. Mok, K. Kim, Phospholips as cancer biomarkers: mass spectrometry-based analysis, Mass Spectrom. Rev. 37 (2018) 107–138, https://doi.org/10.1002/mas.

[33] Y. Bai, H. Zhang, X. Sun, C. Sun, L. Ren, Biomarker identification and pathway analysis by serum metabolomics of childhood acute lymphoblastic leukemia, Clin. Chim. Acta 436C (2014) 207–216, https://doi.org/10.1016/j.cca.2014.05.022.

[34] S.G. Musharraf, A.J. Siddiqui, T. Shamsi, A. Naz, Serum metabolomics of acute lymphoblastic leukaemia and acute myeloid leukaemia for probing biomarker molecules, Hematol. Oncol. 35 (2017) 769–777, https://doi.org/10.1002/hon.2313.

[35] I.C. Kim, J.H. Lee, G. Bang, S.H. Choi, Y.H. Kim, K.P. Kim, H.K. Kim, J. Ro, Lipid profiles for HER2-positive breast cancer, Anticancer Res. 33 (2013) 2467–2472.

[36] S. Ishikawa, I. Tateya, T. Hayasaka, N. Masaki, Y. Takizawa, S. Ohno, T. Kojima, Y. Kitani, M. Kitamura, S. Hirano, M. Setou, J. Ito, Increased expression of phosphatidylcholine (16:0/18:1) and (16:0/18:2) in thyroid papillary cancer, PLoS One 7 (2012), https://doi.org/10.1371/journal.pone.0048873.

[37] N.D. Ridgway, The role of phosphatidylcholine and choline metabolites to cell proliferation and survival, Crit. Rev. Biochem. Mol. Biol. 48 (2013) 20–38, https://doi.org/10.3109/10409238.2012.735643.

[38] Y. Okamoto, A. Aoki, K. Ueda, H. Jinno, Metabolomic analysis uncovered an association of serum phospholipid levels with estrogen-induced mammary tumors in female ACI/Seg rats, Toxicol. Lett. 288 (2018) 65–70, https://doi.org/10.1016/j.toxlet.2018.02.017.

[39] Y. Wang, L. Zhang, W.-L. Chen, J.-H. Wang, N. Li, J.-M. Li, J.-Q. Mi, W.-N. Zhang, Y. Li, S.-F. Wu, J. Jin, Y.-G. Wang, H. Huang, Z. Chen, S.-J. Chen, H. Tang, Rapid diagnosis and prognosis of *de novo* Acute Myeloid Leukemia by serum metabonomic

analysis, J. Proteome Res. 12 (2013) 4393–4401, https://doi.org/10.1021/pr400403p.

[40] Y. Wang, D. Gao, Z. Chen, S. Li, C. Gao, D. Cao, F. Liu, H. Liu, Y. Jiang, Acridone derivative 8a induces oxidative stress-mediated apoptosis in CCRF-CEM leukemia cells: application of metabolomics in mechanistic studies of antitumor agents, PLoS One 8 (2013), https://doi.org/10.1371/journal.pone.0063572.

[41] G.O. Fruhwirth, A. Loidl, A. Hermetter, Oxidized phospholipids: from molecular properties to disease, Biochim. Biophys. Acta (BBA) - Mol. Basis Dis. 1772 (2007) 718–736, https://doi.org/10.1016/j.bbadis.2007.04.009.

[42] T. Nakamura, L. Hall, R.C. Murphy, Oxidation of arachidonate containing glycerophospholipids in intact red blood cells and red blood cell membranes with tert-butylhydroperoxide, in: K. V Honn, L.J. Marnett, S. Nigam, E.A. Dennis (Eds.), Eicosanoids Other Bioact. Lipids Cancer, Inflammation, Radiat. Inj. 4, Springer US, Boston, MA, 1999, pp. 539–545, , https://doi.org/10.1007/978-1-4615-4793-8_79.

[43] S. Tan, X. Liu, T.E. Fox, B.M. Barth, A. Sharma, D. Stephen, A. Awwad, A. Dewey, K. Doi, B. Spitzer, J. Liao, J. Yun, M. Kester, D.F. Claxton, H. Wang, Acid ceramidase is upregulated in AML and represents a novel therapeutic target, Oncotarget 7 (2016) 83208–83222.

[44] L.K. Ryland, T.E. Fox, X. Liu, T.P. Loughran, M. Kester, Dysregulation of sphingolipid metabolism in cancer, Cancer Biol. Ther. 11 (2011) 138–149, https://doi.org/10.4161/cbt.11.2.14624.

[45] G. Cinelli, M. Fabrizi, L. Ravà, M.C.D. Atti, P. Vernocchi, C. Vallone, E. Pietrantoni, R. Lanciotti, F. Signore, M. Manco, Influence of maternal obesity and gestational weight gain on maternal and foetal lipid profile, Nutrients 8 (2016) 1–13, https://doi.org/10.3390/nu8060368.

[46] V.J. Sinanoglou, D. Cavouras, T. Boutsikou, D.D. Briana, D.Z. Lantzouraki, S. Paliatsiou, P. Volaki, S. Bratakos, A. Malamitsi-Puchner, P. Zoumpoulakis, Factors affecting human colostrum fatty acid profile: a case study, PLoS One 12 (2017) 1–14, https://doi.org/10.1371/journal.pone.0175817.

[47] K. Yu, K. Duchen, B. Bjorksten, Fatty acid composition in colostrum and mature breast milk during the first 6 months of lactation, from allergic and non-allergic mothers, Acta Paediatr. 87 (1998) 720–736.

[48] M.A. Mendonca, W.M.C. Araujo, L.A. Borgo, E.D.R. Alencar, Lipid profile of different infant formulas for infants, PLoS One 12 (2017) 1–14, https://doi.org/10.1371/journal.pone.0177812.

[49] M.L. Kwan, P.A. Buffler, B. Abrams, V.A. Kiley, Breastfeeding and the risk of childhood leukemia: a meta-analysis, Publ. Health Rep. 119 (2004) 521–535, https://doi.org/10.1016/j.phr.2004.09.002.

[50] K.R. Greenop, H.D. Bailey, M. Miller, R.J. Scott, J. Attia, L.J. Ashton, P. Downie, B.K. Armstrong, E. Milne, Breastfeeding and nutrition to 2 years of age and risk of childhood acute lymphoblastic leukemia and brain tumors, Nutr. Canc. 67 (2015) 431–441, https://doi.org/10.1080/01635581.2015.998839.

[51] A. Arab, S.A. Akbarian, R. Ghiyasvand, M. Miraghajani, The effects of conjugated linoleic acids on breast cancer: a systematic review, Adv. Biomed. Res. 5 (2016).

[52] K.W. Lee, H.J. Lee, H.Y. Cho, Y.J. Kim, Role of the conjugated linoleic acid in the prevention of cancer, Crit. Rev. Food Sci. Nutr. 45 (2005) 135–144, https://doi.org/10.1080/10408690490911800.

[53] M.K. McGuire, Y. Park, R. Behre, L.Y. Harrison, T.D. Shultz, M. a McGuire, Conjugated linoleic acid concentrations of human milk and infant formula, Nutr. Res. 17 (1997) 1277–1283, https://doi.org/10.1017/CBO9781107415324.004.

[54] A.A. Hennessy, P.R. Ross, G.F. Fitzgerald, C. Stanton, Sources and bioactive properties of conjugated dietary fatty acids, Lipids 51 (2016) 377–397, https://doi.org/10.1007/s11745-016-4135-z.

[55] X.O. Shu, M.S. Linet, M. Steinbuch, W.Q. Wen, J.D. Buckley, P. Joseph, J.D. Potter, H. Gregory, L.L. Robison, Breast-feeding and risk of childhood acute leukemia, J. Natl. Cancer Inst. 91 (1999) 1765–1772.

[56] R.A. Gibson, B. Muhlhausler, M. Makrides, Conversion of linoleic acid and alpha-linolenic acid to long-chain polyunsaturated fatty acids (LCPUFAs), with a focus on pregnancy, lactation and the first 2 years of life, Matern. Child Nutr. 7 (2011) 17–26, https://doi.org/10.1111/j.1740-8709.2011.00299.x.

[57] F. Li, B. He, X. Ma, S. Yu, R.R. Bhave, S.R. Lentz, K. Tan, M.L. Guzman, C. Zhao, H.H. Xue, Prostaglandin E1 and its analog misoprostol inhibit human CML stem cell self-renewal via EP4 receptor activation and repression of AP-1, Cell Stem Cell 21 (2017) 359–373, https://doi.org/10.1016/j.stem.2017.08.001 e5.

[58] Y. Mao, I. Poschke, R. Kiessling, Tumour-induced immune suppression: role of inflammatory mediators released by myelomonocytic cells, J. Intern. Med. 276 (2014) 154–170, https://doi.org/10.1111/joim.12229.