Molecular Cancer

CORRESPONDENCE

Open Access

Epigenome-wide analysis across the development span of pediatric acute lymphoblastic leukemia: backtracking to birth

Akram Ghantous[1*†], Semira Gonseth Nusslé[2†], Farah J. Nassar[1†], Natalia Spitz[1], Alexei Novoloaca[1], Olga Krali[3], Eric Nickels[4,5], Vincent Cahais[1], Cyrille Cuenin[1], Ritu Roy[6], Shaobo Li[4], Maxime Caron[7], Dilys Lam[8,9], Peter Daniel Fransquet[10], John Casement[11], Gordon Strathdee[12], Mark S. Pearce[13], Helen M. Hansen[14], Hwi-Ho Lee[15], Yong Sun Lee[15], Adam J. de Smith[4], Daniel Sinnett[7], Siri Eldevik Håberg[16], Jill A. McKay[17], Jessica Nordlund[3], Per Magnus[16], Terence Dwyer[18,19,20], Richard Saffery[21], Joseph Leo Wiemels[4†], Monica Cheng Munthe-Kaas[16,22†] and Zdenko Herceg[1*†]

Abstract

Background Cancer is the leading cause of disease-related mortality in children. Causes of leukemia, the most common form, are largely unknown. Growing evidence points to an origin *in-utero*, when global redistribution of DNA methylation occurs driving tissue differentiation.

Methods Epigenome-wide DNA methylation was profiled in surrogate (blood) and target (bone marrow) tissues at birth, diagnosis, remission and relapse of pediatric pre-B acute lymphoblastic leukemia (pre-B ALL) patients. Double-blinded analyses was performed between prospective cohorts extending from birth to diagnosis and retrospective studies backtracking from clinical disease to birth. Validation was carried out using independent technologies and populations.

Results The imprinted and immuno-modulating *VTRNA2-1* was hypermethylated (FDR<0.05) at birth in nested cases relative to controls in all tested populations (totaling 317 cases and 483 controls), including European and Hispanic ancestries. *VTRNA2-1* methylation was stable over follow-up years after birth and across surrogate, target and other tissues (*n*=5,023 tissues; 30 types). When profiled in leukemic tissues from two clinical cohorts (totaling 644 cases), *VTRNA2-1* methylation exhibited higher levels at diagnosis relative to controls, it reset back to normal levels at remission, and then re-increased to above control levels at relapse. Hypermethylation was significantly associated with worse pre-B ALL patient survival and with reduced *VTRNA2-1* expression (*n*=2,294 tissues; 26 types), supporting a functional and translational role for *VTRNA2-1* methylation.

†Akram Ghantous, Semira Gonseth Nusslé and Farah J. Nassar equal contribution as first authors.

†Joseph Leo Wiemels, Monica Cheng Munthe-kaas and Zdenko Herceg equal contribution as last authors.

*Correspondence:
Akram Ghantous
GhantousA@iarc.who.int
Zdenko Herceg
HercegZ@iarc.who.int
Full list of author information is available at the end of the article

Conclusion This study provides proof-of-concept to detect at birth epigenetic precursors of pediatric pre-B ALL. These alterations were reproducible with different technologies, in three continents and in two ethnicities, and can offer biomarkers for early detection and prognosis as well as actionable targets for therapy.

Key points

- Precursors of pediatric acute lymphoblastic leukemia may be of epigenetic origin, detectable since birth and affecting patient prognosis.
- These epigenetic precursors can be robust over several years and across several populations, ethnicities and surrogate and target tissues.

Keywords Pediatric leukemia, Epigenetics, DNA methylation, *VTRNA2-1*, Birth cohort, Neonatal blood spots

Background

Pediatric acute lymphoblastic leukemia (ALL) is the most common childhood cancer [1] and originates either from mature B-cells (2%), T-cells (15%), or early/precursor B-cells (80-85% of ALL) [2]. Despite high ALL survival, efforts towards its prevention are warranted due to relapse and long-term adverse effects of therapy [2].

Unlike adult leukemia, the majority of pediatric leukemias do not have well-established causes. Because ALL is rare, the limited evidence on risk factors stems mostly from retrospective studies, offering abundant samples with limitations of recall and selection bias [3]. Moreover, studies on underlying molecular causes are based largely on biospecimen collected after disease onset, with likelihood of reverse causality bias, wherein identified molecular alterations can be due to cancer rather than its cause [3, 4]. Prospective designs would be ideal, but no single cohort can collect enough pre-diagnostic biospecimens given ALL rarity. Hence, international efforts and novel approaches are crucial.

Compared to adult cancers, the overall mutation burden is generally low in pediatric leukemias, and the most commonly altered genes are epigenetic regulators [5], associated with DNA methylome-wide alterations [6, 7]. Thus, it is plausible that epigenetic mechanisms play a central role in pediatric cancer development particularly that it may have an origin *in-utero* [4, 8, 9], a period of cellular programming driven largely by epigenetics.

Hence, we sought to identify for the first time genome-wide methylation alterations in newborns before the onset of pediatric precursor B-cell ALL (pre-B ALL) and assess the translational potential of the findings in pre- and post-diagnosis periods using different technologies, populations, ethnicities and tissue matrices, including surrogate and target tissues.

Results and discussion

DNA methylation alterations in neonatal blood associated with pre-B ALL development

Epigenome-wide analysis in the prospective MoBa (Norway) and the retrospective CCLS (USA) discovery datasets (Supplementary Table 1) identified significant (FDR<0.05) differentially methylated regions (DMRs) in the blood of newborns who later developed pediatric pre-B ALL, relative to controls (Supplementary Fig. 1; Supplementary Tables 2A-B). DMRs in both studies were significantly enriched in CpG Shores, Promoters, First Exons, Exon-Intron/Intron-Exon boundaries and imprinted genes while being depleted in Open Sea, Shelf and intronic regions (p<0.05) (Supplementary Fig. 2). Both studies significantly converged (p<0.01) on several CpGs (Supplementary Tables 2A-B), among which 7 passed all filters (Fig. 1A), including effect size ≥ 3% (considered sufficiently high for validation by targeted sequencing) (Supplementary Fig. 1). All 7 CpGs mapped to the imprinted *VTRNA2-1*, which showed hypermethylation in nested cases relative to controls (Fig. 1A). *VTRNA2-1* encompassed 9 additional CpGs, which were significant (FDR<0.05) in both studies' Crude Models (Supplementary Table 2A). *VTRNA2-1* methylation showed sex-dependent but ethnicity-independent alterations in the European and Hispanic descents (Fig. 1B), with the effects being observed in females, though requiring validation in larger sample sizes given that stratification by sex reduces power. In the replication phase, findings were reproducible in independent samples from MEDC (Australia) using a different technology, EpiTyper (Fig. 1C).

Functional analysis of *VTRNA2-1*

As expected for an imprinted gene, *VTRNA2-1* methylation levels were centered around 50±10% and were similar across various tissue types, including bone marrow (pre-B ALL target tissue) and cord blood (surrogate tissue)

(p>0.05), with the exception for placenta (being extra-embryonal) and sperm (being hypomethylated given that *VTRNA2-1* is maternally imprinted) (Fig. 1D). In various tissue types, *VTRNA2-1* methylation negatively correlated with its gene transcription (FDR<0.05), including in pediatric pre-B ALL tissues (Supplementary Fig. 3 and Fig. 1E).

Longitudinal analysis in controls and cases pre-diagnosis

Assessment of *VTRNA2-1* methylation stability over time showed that methylation at all 16 CpGs was similar across birth years in cord blood samples collected from controls and pre-diagnostic cases from 2000-2008 in MoBa (Supplementary Fig. 4A), and this was replicated in neonatal blood spots collected from 1984-1994 in an independent population, UKCS (UK) (Supplementary Fig. 4B). Moreover, there was no significant difference (p>0.05) in *VTRNA2-1* methylation levels between cord blood (i.e. at birth) and peripheral blood collected from the same individuals (controls) at age three years (Fig. 1F), which is the peak incidence age of pre-B ALL [12]. This highlights the stability of *VTRNA2-1* methylation over critical time windows and across matched cord and peripheral blood tissues, reinforcing its observed methylation stability in a panel of human tissues (Fig. 1D).

Longitudinal analysis post-diagnosis

In line with the *VTRNA2-1* hypermethylation observed at birth in nested cases *versus* controls, *VTRNA2-1* was significantly hypermethylated in pediatric pre-B ALL tissues at diagnosis compared to controls, regardless of diagnostic matrix (surrogate=blood or target=bone marrow, Fig. 2A). This was based on the NOPHO (Nordic countries) cohort and validated in the QcALL cohort (Canada) (Supplementary Fig. 5). At remission, *VTRNA2-1* methylation levels were reset to normal, and, at relapse, they re-increased to above control levels (Fig. 2A). This trend was validated in subjects matched at diagnosis and remission in QcALL (Fig. 2B). Overall, these results suggest *VTRNA2-1* methylation as a marker of pre-B ALL prognosis, including leukemic state.

(See figure on next page.)

Fig. 1 Discovery, validation and functional analysis of *VTRNA2-1* methylation in association with pediatric pre-B ALL development. **A** Upper section: Prioritized differentially methylated genes with at least one CpG with effect size ≥3% after DMR analysis for blood samples taken from newborns of either MoBa or CCLS. 7 CpG sites were significantly enriched ($p = 2.2 \times 10^{-16}$) between MoBa and CCLS relative to the total number of array CpGs analyzed (470,963 CpGs); all CpGs mapped to the same gene, which was also significantly enriched ($p = 4.4 \times 10^{-3}$) relative to the total number of genes in the human genome (21,306 genes) (Fisher's Exact Test). Lower section: The 7 significant CpGs within the DMR of *VTRNA2-1* are symbolized in CCLS and MoBa by circles of varying sizes and colors, representing the effect sizes and directions of effect, respectively, as per the figure legend. The 7 CpGs are arranged in order according to their genomic position. The direction of effect is reported for the pre-B ALL nested cases relative to the controls: hypermethylation (Hypermeth) or hypomethylation (Hypometh). **B** *VTRNA2-1* differential methylation in nested cases and controls was stratified by subject sex and ethnicity in MoBa and CCLS cohorts. Data points represent average methylation values at each CpG site, and the ribbons denote the 95% confidence intervals. CpG HM450 IDs are shown on the x-axes. In addition to the CpGs (in red) identified in the Adjusted Models in both MoBa and CCLS, we also show (in black) the additional CpGs identified in the Crude Models in both MoBa and CCLS (detailed in Supplementary Fig. 7 and Supplementary Table 2). **C** Validation, based on profiling of *VTRNA2-1* methylation using EpiTyper, which is sequencing- rather than array-based, applied to an independent set of biological samples from MEDC. Data points represent average methylation values at each CpG site, and the error bars denote the 95% confidence intervals. The *p*-values indicate the statistical significance across the whole DMR region and were calculated by inverse variance based meta-analysis using METAL software. The DMR profiled by EpiTyper partially overlaps with that by HM450; specifically, CpGs 10 and 11 in (**C**) are identical to the last two CpGs in (**B**), cg16615357 and cg18797653, respectively. CpG1-2 and CpG3-4 each represents an average methylation value of two adjacent CpGs, as detected by EpiTyper. The genomic coordinates of the CpG ID numbers are detailed in Supplementary Fig. 8. **D** Box plots showing the methylation distribution of *VTRNA2-1* across a panel of human tissue types using data extracted from the EWAS Open Platform [10]. The box plots encompass the first quartile (bottom border), the median (middle line), the fourth quartile (upper border) and the extreme values (dots). No statistically significant differences (p>0.05; Mann-Whitney test) were detected in *VTRNA2-1* mean methylation between the target bone marrow and surrogate cord blood tissues. The sample sizes (N) are indicated for each tissue type. (n=5,023 tissues; 30 types) **E** Pearson correlation of *VTRNA2-1* expression with the methylation of its CpGs in a panel of cancer tissues extracted from the MEXPRESS database [11] (n=2,273 tissues; 25 types). Cancer types and sample sizes are as follows: kidney renal papillary cell carcinoma (KIRP, N = 140), rectum adenocarcinoma (READ, N = 28), pheochromocytoma and paraganglioma (PCPG, N = 66), skin cutaneous carcinoma (SKCM, N = 74), testicular germ cell tumor (TGCT, N = 47), uveal melanoma (UVM, N = 28), thyroid carcinoma (THCA, N = 162) , kidney renal clear cell carcinoma (KIRC, N = 136), breast invasive carcinoma (BRCA, N = 106), pancreatic adenocarcinoma (PAAD, N = 67), colon adenocarcinoma (COAD, N = 95), prostate adenocarcinoma (PRAD, N = 75), liver hepatocellular carcinoma (LIHC, N = 90), bladder urothelial (BLCA, N = 106), uterine corpus endometrial carcinoma (UCEC, N = 121), head and neck squamous cell carcinoma (HNSC, N = 103), lung adenocarcinoma (LUAD, N = 74), mesothelium (MESO, N = 42), lung squamous cell carcinoma (LUSC, N = 44), glioblastoma multiforme (GBM, N = 44), sarcoma (SARC, N = 82), cervical squamous cell carcinoma and endocervical adenocarcinoma (CESC, N = 127), brain lower grade glioma (LGG, N = 96), stomach adenocarcinoma (STAD, N = 217) and esophageal carcinoma (ESCA, N = 103). The asterisk (*) mark significant correlation after adjustment for multiple testing (FDR < 0.05). One CpG (cg11978884) was omitted from the analysis because it had no methylation values. **F** *VTRNA2-1* methylation in MoBa paired samples over time. None of the *VTRNA2-1* CpGs were significantly (p>0.05) differentially methylated in cord blood collected from the control subjects at age 0 (blue) *versus* paired peripheral blood collected from the same controls at age 3 (orange) years (Wilcoxon test). Methylation values at birth from nested unpaired controls (green) and cases (red) are shown as a reference. The y-axes represent the methylation (beta) values, and p values are reported for each CpG. In E-F, the orange rectangles represent CpGs common to Adjusted Models of both MoBa and CCLS. The remaining CpGs are those identified in the Crude Models of both datasets

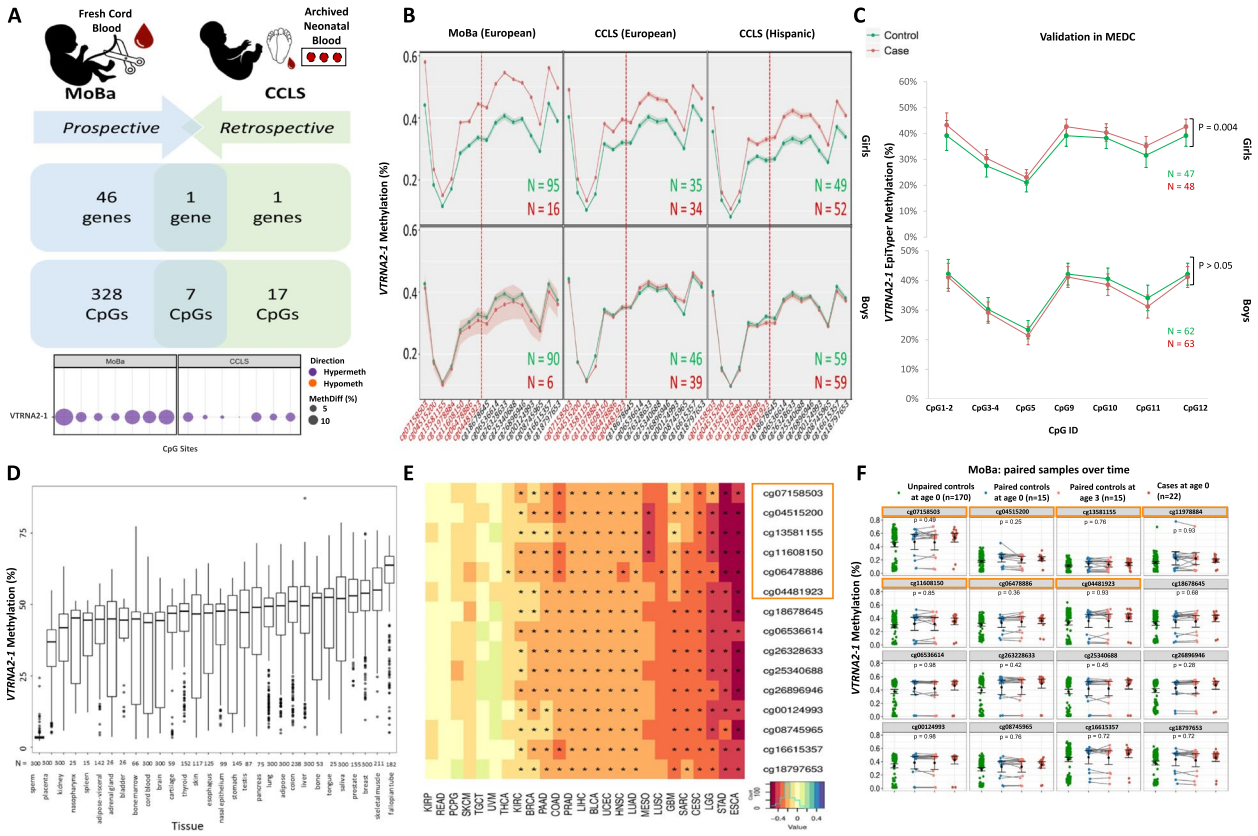

Fig. 1 (See legend on previous page.)

Focusing on individual-level data, we observed two QcALL patient clusters: C1 exhibited low methylation levels (10±10%), stable from diagnosis to remission (p>0.05), and C2 exhibited at diagnosis higher methylation levels (~20-100%), which converged to 50±10% at remission (p<0.05) (Fig. 2B and Supplementary Fig. 6A-B). The C1 and C2 patterns were also observed in blood at birth (Supplementary Fig. 6C) and in NOPHO (Supplementary Fig. 6D) and are in line with recent observations showing that *VTRNA2-1* is imprinted in

(See figure on next page.)

Fig. 2 Longitudinal analysis of *VTRNA2-1* methylation post-diagnosis and its hypothesized role in pre-B ALL development. **A** Methylation of *VTRNA2-1* CpGs in peripheral blood and bone marrow of cases *versus* controls at diagnosis, remission and relapse in NOPHO. In purple: Methylation of *VTRNA2-1* CpGs in peripheral blood samples from sorted B-cells of normal subjects (*N*=26) and from pediatric pre-B ALL patients collected at diagnosis (*n*=74) from NOPHO. In green: Methylation of *VTRNA2-1* CpGs in sorted B-cells (*N*=26) from bone marrow of fetuses (*N*=8) and in bone marrow samples from cases of pediatric pre-B ALL collected at diagnosis (*n*=535), remission (*n*=82) and relapse (*n*=32) from NOPHO. Whiskers represent the minimum and the maximum, while the top, the bottom, and the band in the box represent the first and third quartile and the median respectively. Significant differences between methylation of normal and tumor samples are marked for each CpG with an asterisk (Wilcoxon test). **B** Methylation of *VTRNA2-1* CpGs in 46 pediatric pre-B ALL samples collected at diagnosis (red) and remission (blue) from the same patients in QcALL. Significant differences between methylation at diagnosis and remission are marked for each CpG with an asterisk (Wilcoxon test). The data are represented in the form of a dot plot to better visualize the paired samples (a line links each pair). Red and blue box plots are also shown for each time point (diagnosis and remission, respectively), representing the first quartile (bottom border), the median (middle line) and the fourth quartile (upper border) for each condition. **C** and **D** Methylation of *VTRNA2-1* CpGs in relation to overall and relapse-free survival, respectively, represented by hazard ratios. In NOPHO, 598 pre-B ALL patients were followed up for ten years or more. *VTRNA2-1* methylation at two CpG sites significantly affected overall survival (denoted by *, Wald test), after adjusting for patient sex, age and risk groups using a Multivariate Cox Regression model. Risk group variables also affected overall and relapse-free survival (denoted by ** or ***, Wald test). HR: high risk, IR: intermediate risk and SR: standard risk. * denotes p<0.05, ** denotes p<0.01, *** denotes p<0.001. In A-D, the orange rectangles represent CpGs common to Adjusted Models of both MoBa and CCLS. The remaining CpGs are those identified in the Crude Models of both datasets. **E** The tumor surveillance model offering a biologically plausible mechanism of *VTRNA2-1* in pediatric pre-B ALL development. The basal methylation and expression level of *VTRNA2-1* determines the degree of gradients (narrow: RIGHT *versus* wide: LEFT), which is important to shift the balance from cell survival (RIGHT) to cell death (LEFT) *via* PKR activation. Graphic icons used to construct the figure were retrieved from thenounproject.com. **F** Summary of the study's time points, sample types, and *VTRNA2-1* results

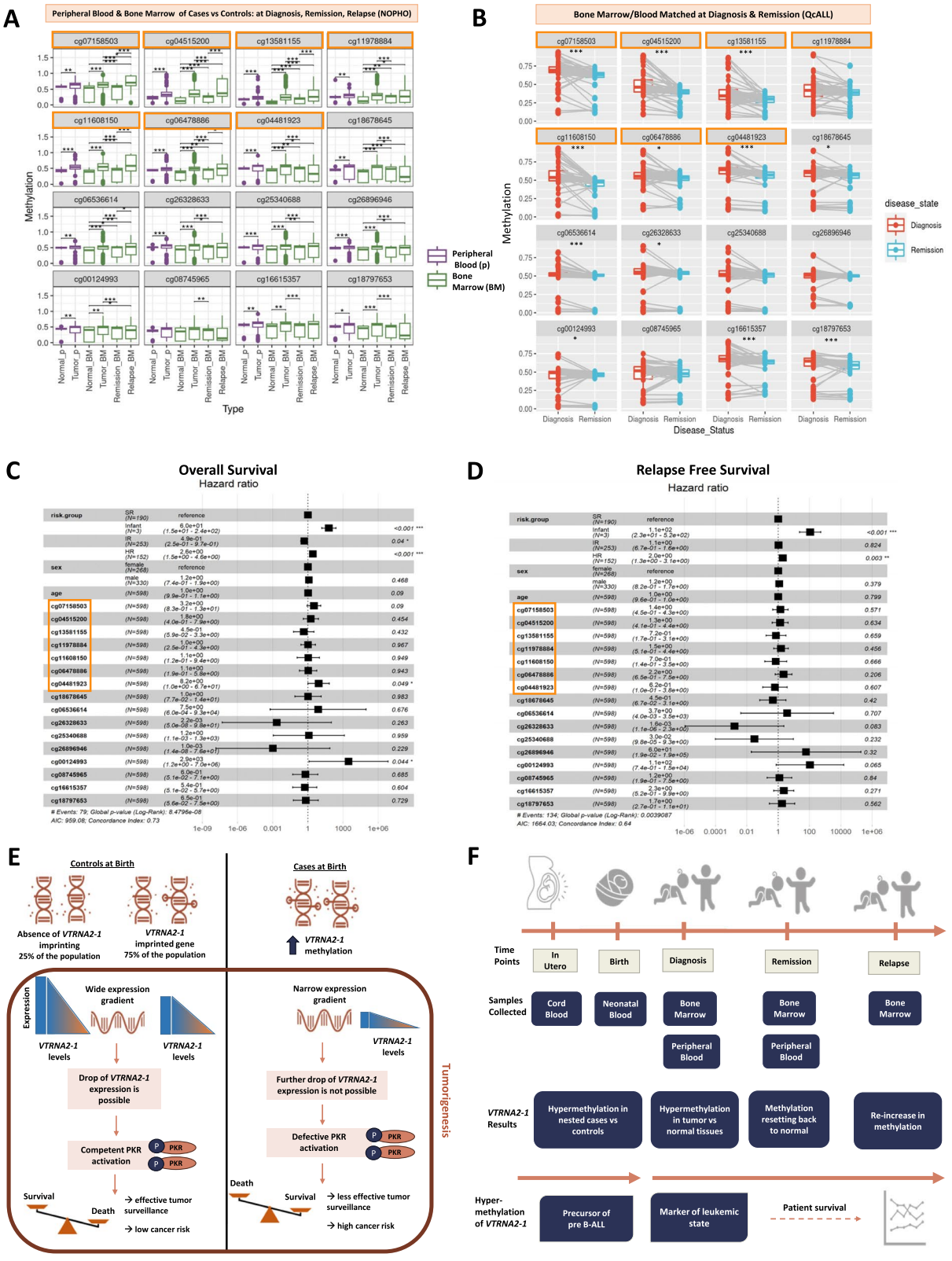

Fig. 2 (See legend on previous page.)

~75% of individuals [13], as in C2, and non-imprinted in the remaining portion (C1).

To test whether *VTRNA2-1* methylation affects clinical outcomes, ten-year follow-up data was used in NOPHO (n=598, including 134 relapse and 79 death events; Supplementary Table 3). *VTRNA2-1* methylation at two CpG sites was significantly inversely associated with overall survival (hazard ratio>1; p<0.05), after adjusting for sex, age and prognosis risk groups (Fig. 2C), hence, reinforcing the prognostic potential of this gene. No significant difference (p>0.05) on patient overall or relapse-free survival was detected between C1 and C2 (Supplementary Fig. 6E-F). Also, no significant associations (p>0.05) were observed between *VTRNA2-1* methylation and relapse-free survival (Fig. 2D), although relapse events were more frequent (exhibiting higher statistical power) than death events, suggesting that *VTRNA2-1* methylation may likely associate with patient survival more than relapse.

Hypothetical model of *VTRNA2-1* role pre- and post-diagnosis

VTRNA2-1 is a 100-nucleotide non-coding RNA with central roles in multiple cancer types based on cell, animal and clinical models and is a known regulator of Protein Kinase R (PKR)-mediated cell death (Supplementary Table 4). A tumor surveillance model for eliminating precancerous cells has been proposed requiring *VTRNA2-1* hypermethylation and the 'drop' in its expression as a critical event for PKR activation [14] (Fig. 2E). The effectiveness of this mechanism is speculated to be weaker in individuals in which *VTRNA2-1* expression is already low (i.e. hypermethylated) at birth, prohibiting any possible drop of *VTRNA2-1* levels and subsequent cell death. This can lead since birth to the accumulation of precancerous cells that can malignantly transform over time (Fig. 2E). *VTRNA2*-1 methylation could also affect PKR-mediated immune regulation [15], which can serve as an additional hit to activate pre-leukemic clones to progress into malignancy and/or could enable existing tumor cells to evade immune attack (hence, worsening prognosis).

Conclusions

This work represents a proof-of-concept to detect at birth epigenetic precursors of pediatric pre-B ALL by mapping for the first time the epigenome across the development span of pediatric leukemia (*in utero*, birth, diagnosis, remission and relapse) (Fig. 2F) and liaising prospective studies operating from birth to diagnosis with retrospective studies backtracking from clinical disease to birth (Supplementary Fig. 1). As such, sample sizes of rare biospecimen are augmented (through retrospective studies) while recall, selection and reverse causality biases are reduced (through prospectively collected data and pre-diagnostic biospecimen), therefore, strengthening causality of associations particularly when implemented in various populations, which offer a natural means of effect randomization. Among identified significant genes, *VTRNA2-1* methylation alterations in neonatal blood were reproducible with different technologies, in three continents and in two ethnicities. Epigenetic alterations detectable before diagnosis could serve as biomarkers for early detection and as precursors of pediatric B-ALL rather than resultant passengers. Blood-based biomarkers are easy to measure and amenable to population screening especially using cost-effective targeted sequencing such as EpiTyper. Moreover, *VTRNA2-1* methylation was consistent across various tissue types and showed prognostic potential linked to leukemic state and patient survival. This, along with its possible precursor role and its association with prognosis in several cancer types (Supplementary Table 4), makes *VTRNA2-1* a promising target for epigenetic therapy.

Future work based on more cases may be able to uncover further molecular precursors of pediatric pre-B ALL especially in relation to various age groups, chromosomal aberrations, ethnicities and subject sex. This work addresses a timely need of ethnic diversification in research studies by including European and underrepresented USA Hispanic children, who have the highest pediatric leukemia rates worldwide [1]. The identification of an epigenetic signature at birth associated with the risk and the prognosis of pediatric B-ALL may change our paradigm of leukemogenesis by uncovering molecular origins of leukemia since the time of birth.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12943-024-02118-4.

Supplementary Material 1.

Supplementary Material 2.

Supplementary Material 3.

Acknowledgements

We are grateful to Dr. Karen Müller for assistance with editing this manuscript for language and grammar. We are grateful to all the participating families in Norway who participate in the ongoing MoBa cohort study. For recruitment of subjects enrolled in CCLS, the authors gratefully acknowledge the clinical investigators at the following collaborating hospitals: 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 (formerly Sacramento) (Drs. Kent Jolly and Vincent Kiley), Kaiser Permanente Santa Clara (Drs. Carolyn Russo, Alan Wong, and Denah Taggart), Kaiser Permanente San Francisco (Dr. Kenneth Leung), and Kaiser Permanente Oakland (Drs. Daniel Kronish and Stacy Month). The authors thank the Newcastle Central Biobank Facility for curation and use of neonatal blood spot samples, Dr Richard McNally and Mr Richard Hardy for linkage to the Northern Region Young Persons' Malignant Disease Registry, Dr Jessica Timms for assistance

with sample identification and Genetics Core, Edinburgh Clinical Research Facility, University of Edinburgh for sample processing. We thank the I4C International Data Coordinating Center in Melbourne, Australia (in particular, Dr. Gabriella Tikellis) for assistance in data management in I4C cohorts. We thank Dr. Tomi Pastinen for his assistance in data management of the QcALL DNA methylome dataset. Sincerest thanks to the participating children and their families to whom we dedicate this manuscript.

Disclaimer

Where authors are identified as personnel of the International Agency for Research on Cancer/World Health Organization, the authors alone are responsible for the views expressed in this article and they do not necessarily represent the decisions, policy or views of the International Agency for Research on Cancer/World Health Organization. The California Department of Public Health is not responsible for the results or conclusions drawn by the authors of this publication.

Authors' contributions

Conceptualization: AG, JLW, MCMK, TD and ZH. Data collection and/or generation: AJDS, CC, DL, DS, FJN, HMH, JLW, JM, JN, MCMK, PDF, RS and SGN. Statistical design of sample selection: AG, GS, JM, MCMK, MP, PM, SEH and TD. Data analysis: AG, AN, DL, FJN, JC, JM, MC, NS, OK, RR, SGN, SL and VC. Data interpretation: AG, AN, DS, FJN, H-HL, JLW, JN, MC, MCMK, NS, OK, PM, RS, SEH, SGN, TD, YSL and ZH. Funding: AG, JLW, JM, TD and ZH. Supervision: AG, JLW, JN, RS and ZH. Writing original draft: AG. Editing or reviewing the manuscript: all authors. All authors approved the final version of the manuscript.

Funding

This work was supported by the National Cancer Institute Pediatric Oncology High Risk-High Gain grant (PEDIAHRG-2020, INCa_15817, France, 2021-2024, PI: AG), INCa/Plan Cancer-EVA-INSERM (France, 2015-2019, PI: ZH) and INCa (PEDIAC INCa_15670, PI: ZH). AG had also been supported by the IARC Postdoctoral Fellowship and Marie Curie Actions-People-COFUND (2012-2014). NS was supported by the IARC Postdoctoral Fellowship at the International Agency for Research on Cancer, financed by Children with Cancer UK (2021-2022). SG was supported by a grant of the Swiss National Foundation (P2LAP3_158674 / 3, 2015-2016) and by the SICPA Foundation. OK and JN are supported by the Swedish Research Council (2019-01976), the Swedish Childhood Cancer Foundation (PR2019-0046), and the Göran Gustafsson Foundation.

MoBa is supported by the Norwegian Ministry of Health and Care Services and the Ministry of Education and Research. The work was partly funded by the Norwegian Research Council's Centre of Excellence funding scheme, through "Centre for Fertility and Health" (project no. 262700).

The CCLS work was supported by research grants from the National Institutes of Health (R01CA175737, PI: JLW; R01ES009137; and P01ES018172, PI: JLW, Project 3) and the Environmental Protection Agency (RD83451101, PI: JLW, Project 3), USA. The California-based biospecimens and/or data used in this study were obtained from the California Biobank Program, (SIS request #26), Section 6555 (b), 17 CCR.

The UKCS study was funded by The JGW Patterson Foundation (BH142153, PI: JM) with salary support for JM from Children's Cancer North.

Data availability

Data from MoBa and the Medical Birth Registry of Norway used in this study are managed by the national health register holders in Norway (Norwegian Institute of Public Health) and can be made available to researchers, provided there is approval from the Regional Committees for Medical and Health Research Ethics (REC), compliance with the EU General Data Protection Regulation (GDPR) and approval from the data owners. The consent given by the participants does not open for storage of data on an individual level in repositories or journals. Researchers who want access to data sets for replication should apply through helsedata.no. Access to data sets requires approval from The Regional Committee for Medical and Health Research Ethics in Norway and an agreement with MoBa.

CCLS data is derived from the California Biobank. We respectfully are unable to share raw, individual level data freely with other investigators since the samples and the data are the property of the State of California. Should we be contacted by other investigators who would like to use the data, we will direct them to the California Department of Public Health Institutional Review Board to establish their own approved protocol to utilize the data, which can then be shared peer-to-peer. The State has provided guidance on data sharing noted in the following statement: "California has determined that researchers requesting the use of California Biobank biospecimens for their studies will need to seek an exemption from NIH or other granting or funder requirements regarding the uploading of study results into an external bank or repository (including into the NIH dbGaP or other bank or repository). This applies to any uploading of genomic data and/or sharing of these biospecimens or individual data derived from these biospecimens. Such activities have been determined to violate the statutory scheme at California Health and Safety Code Section 124980 (j), 124991 (b), (g), (h) and 103850 (a) and (d), which protect the confidential nature of biospecimens, and individual data derived from biospecimens. All investigators seeking to use California specimens for projects or grant related activities that require or seek such sharing (at the NIH or elsewhere) must seek an exemption from genomic data sharing requirements. If such an exemption is not secured, samples will not be released to an investigator."

For QcALL, DNA methylation (HM450) data has been deposited in the GEO database under accession code GSE38235, and RNA-seq data is available for of 21 out of 46 diagnosis patients under accession code GSE89071.

For UKCS, DNA methylation (HM850) data has been deposited in the GEO database under accession code GSE211591.

For NOPHO, DNA methylation (HM450) data has been deposited in the GEO database under accession code GSE49031.

For sorted cells from fetuses, DNA methylation (HM450) data is available in the GEO database under GSE45459.

For EWAS Open Platform [10], DNA methylation (HM450) across a panel of human tissue types is available at https://ngdc.cncb.ac.cn/ewas/datahub/download.

For TCGA, RNA expression and DNA methylation data across a panel of cancer types were accessible and analyzed through MEXPRESS database [11].

Declarations

Ethics approval and consent to participate

The project was approved by the IARC Ethics Committee. The establishment of MoBa and initial data collection was based on a license from the Norwegian Data Protection Agency and approval from The Regional Committees for Medical and Health Research Ethics. The MoBa cohort is currently regulated by the Norwegian Health Registry Act. The current study was approved by The Regional Committees for Medical and Health Research Ethics and the Norwegian Data Inspectorate (REK number 2011/2551b), and written informed consent was provided by both parents.

For CCLS, written informed consent was obtained by the legal representative of all included participants. The California State and University of California IRBs have approved the study.

For MEDC, newborn blood spot punches from children later diagnosed with ALL were obtained from archival Guthrie cards under a waiver of consent granted as part of the project 'Epigenetic determinants of Childhood Cancer Incidence' Human Research Ethics Committee Approval (HRE/16/RCHM/62).

The UKCS study used samples from the Newcastle Biobank which holds approvals from the Newcastle and North Tyneside Research Ethics Committee 1 (17/NE/0361). Additional approvals specific to the use of neonatal blood spot samples and linkage of these samples to the Northern Region Young Persons' Malignant Disease Registry were obtained from the Health Research Authority Confidentiality Advisory Group (17CAG0177) and Newcastle and North Tyneside Research Ethics Committee 1 (17/NE/0334).

For QcALL, informed consent was obtained from all participating individuals, and the Sainte-Justine UHC Institutional Review Board approved the research protocols.

For NOPHO, the study was approved by the Regional Ethical Review Board in Uppsala, Sweden and was conducted according to the guidelines of the Declaration of Helsinki. The patients and/or their guardians provided informed consent.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details
[1]Epigenomics and Mechanisms Branch, International Agency for Research on Cancer, (IARC), 25 avenue Tony Garnier, CS 90627, Lyon Cedex 07 69366, France. [2]Genknowme SA, Alanine Building, Corniche Road 5, 1066 Epalinges, Switzerland. [3]Department of Medical Sciences, Molecular Precision Medicine and Science for Life Laboratory, Uppsala University, Uppsala, Sweden. [4]Center for Genetic Epidemiology, University of Southern California, Los Angeles, USA. [5]Cancer and Blood Disease Institute, Children's Hospital Los Angeles, Los Angeles, USA. [6]Computational Biology and Informatics Core, Helen Diller Family Comprehensive Cancer Center, University of California San Francisco, San Francisco, USA. [7]Department of Pediatrics, University of Montreal & Research Center, Centre Hospitalier Universitaire (CHU) Sainte-Justine, Montréal, Quebec, Canada. [8]School of Molecular Sciences, The University of Western Australia, Crawley, Western Australia 6009, Australia. [9]Harry Perkins Institute of Medical Research, Nedlands, Western Australia 6009, Australia. [10]Centre for Social and Early Emotional Development (SEED), School of Psychology, Deakin University, Burwood, Victoria, Australia. [11]Bioinformatics Support Unit, Newcastle University, Newcastle upon Tyne, UK. [12]Newcastle University Centre for Cancer, Biosciences Institute, Newcastle University, Newcastle upon Tyne, UK. [13]Population Health Sciences Institute, Newcastle University, Newcastle upon Tyne, UK. [14]Neuro and Molecular Epidemiology Laboratory, University of California San Francisco, San Francisco, USA. [15]Department of Cancer Biomedical Science, Graduate School of Cancer Science and Policy, National Cancer Center, Goyang 10408, Korea. [16]Centre for Fertility and Health, Norwegian Institute of Public Health, Oslo, Norway. [17]Department of Applied Sciences, Northumbria University, Newcastle upon Tyne, UK. [18]Clinical Sciences Theme, Heart Group, Murdoch Children's Research Institute, Melbourne, Australia. [19]Nuffield Department of Women's & Reproductive Health, University of Oxford, Oxford, UK. [20]Faculty of Medicine, Dentistry and Health Sciences, University of Melbourne, Melbourne, Australia. [21]Murdoch Children's Research Institute and Department of Paediatrics, University of Melbourne, Melbourne, Australia. [22]Department of Pediatric Oncology and Hematology, Oslo University Hospital, Oslo, Norway.

Received: 7 June 2024 Accepted: 9 September 2024
Published online: 23 October 2024

References

1. Steliarova-Foucher E, Colombet M, Ries LAG, Moreno F, Dolya A, Bray F, et al. International incidence of childhood cancer, 2001–10: a population-based registry study. Lancet Oncol. 2017;18(6):719–31.
2. Hunger SP, Mullighan CG. Acute lymphoblastic leukemia in children. N Engl J Med. 2015;373(16):1541–52.
3. Erdmann F, Ghantous A, Schüz J. Environmental agents and childhood cancer. In: Nriagu J, editor. Encyclopedia of Environmental Health. 2nd ed. Burlington: Elsevier B.V.; 2019. p. 336–47. https://scholar.google.com/scholar_lookup?title=Encyclopedia+of+Environmental+Health&author=F+Erdmann&author=A+Ghantous&author=J+Sch%C3%BCz&author=J+Nriagu&publication_year=2019&.
4. Ghantous A, Hernandez-Vargas H, Byrnes G, Dwyer T, Herceg Z. Characterising the epigenome as a key component of the fetal exposome in evaluating in utero exposures and childhood cancer risk. Mutagenesis. 2015;30(6):733–42.
5. Gröbner SN, Worst BC, Weischenfeldt J, Buchhalter I, Kleinheinz K, Rudneva VA, et al. The landscape of genomic alterations across childhood cancers. Nature. 2018;555(7696):321–7.
6. Nordlund J, Syvänen AC. Epigenetics in pediatric acute lymphoblastic leukemia. Semin Cancer Biol. 2018;51:129–38.
7. Nordlund J, Bäcklin CL, Wahlberg P, Busche S, Berglund EC, Eloranta ML, et al. Genome-wide signatures of differential DNA methylation in pediatric acute lymphoblastic leukemia. Genome Biol. 2013;14(9):r105.
8. Wiemels J. Perspectives on the causes of childhood leukemia. Chem Biol Interact. 2012;196(3):59–67.
9. Greaves M. A causal mechanism for childhood acute lymphoblastic leukaemia. Nat Rev Cancer. 2018;18(8):471–84.
10. Xiong Z, Yang F, Li M, Ma Y, Zhao W, Wang G, et al. EWAS Open Platform: integrated data, knowledge and toolkit for epigenome-wide association study. Nucleic Acids Res. 2022;50(D1):D1004–9.
11. Koch A, Jeschke J, Van Criekinge W, van Engeland M, De Meyer T. MEXPRESS update 2019. Nucleic Acids Res. 2019;47(W1):W561–5.
12. Hjalgrim LL, Rostgaard K, Schmiegelow K, Söderhäll S, Kolmannskog S, Vettenranta K, et al. Age- and sex-specific incidence of childhood leukemia by immunophenotype in the Nordic countries. JNCI. 2003;95(20):1539–44.
13. Carpenter BL, Zhou W, Madaj Z, DeWitt AK, Ross JP, Grønbæk K, et al. Mother-child transmission of epigenetic information by tunable polymorphic imprinting. Proc Natl Acad Sci USA. 2018;115(51):E11970–7.
14. Jeon SH, Johnson BH, Lee YS. A tumor surveillance model: a non-coding RNA senses neoplastic cells and its protein partner signals cell death. Int J Mol Sci. 2012;13(10):13134–9.
15. Lee YS, Bao X, Lee HH, Jang JJ, Saruuldalai E, Park G, et al. Nc886, a novel suppressor of the type I interferon response upon pathogen intrusion. Int J Mol Sci. 2021;22(4):2003.

Publisher's Note