

LETTER OPEN



GENOMICS AND GENE THERAPY

A DNA methylation database of human and mouse hematological malignancy cell lines

Aleix Noguera-Castells ^{1,2,3}, Carlos A. García-Prieto ^{1,4}, Gerardo Ferrer ^{1,2}, Veronica Davalos^{1,2}, Fernando Setien¹, Eulàlia Genescà ¹, Jordi Ribera¹, Josep M. Ribera ^{1,5,6} and Manel Esteller ^{1,2,7,8}✉

© The Author(s) 2024

Leukemia (2025) 39:512–515; <https://doi.org/10.1038/s41375-024-02478-2>

TO THE EDITOR:

Human cancer cell lines constitute useful models to study the primary disease and many relevant findings in tumor biology have been originated from them since the first immortalized cell line (HeLa) was obtained [1]. The easy experimental intervention, the purity of the transformed cells and their versatility to undergo high-throughput screenings represent advantageous features of the established cancer cell lines. In recent years, major efforts have characterized in detail the multiomics make-up of hundreds of cancer cell lines and studied their association with sensitivity to anticancer drugs [2–5]. Most of these studies have been genetic-centric and have not characterized in detail the epigenetic profiles underlying the characteristics of these cells or their impact on the efficacy of antitumoral compounds. In this regard, a past version of a DNA methylation microarray [3, 6, 7] or a more time-consuming readout such as reduced representation bisulfite sequencing [4, 8] have been used in those attempts to interrogate the epigenetic setting. Herein, we have obtained the DNA methylation profiles of 210 cell lines derived from hematological malignancies utilizing comprehensive DNA methylation microarrays that interrogates more than 850,000 and 285,000 CpG sites from human and mouse genomes, respectively [9, 10]. Importantly, we also provide a pharmacoepigenetic example of the potential use of this resource by showing how DNA methylation can predict response to nucleoside analogues.

In this regard, using the above-described platforms [9, 10], we analyzed the DNA methylation profile of 180 and 30 human and mouse hematological cell lines, respectively, as illustrated in Fig. 1A (Supplementary Methods). Overall, these 210 samples encompassed 80 lymphoma, 93 leukemia, 20 multiple myeloma and 17 non-malignant transformed cell types (Fig. 1A). The entire description of all the studied cell lines is shown in Dataset S1. The complete DNA methylation data are freely available at the GEO repository under accession number GSE270494. For both human and mouse, we found by using the whole DNA methylome that

the hematological cell lines tend to cluster by disease (leukemia, lymphoma or multiple myeloma) as shown in the unsupervised hierarchical clustering (Fig. 1B) (Supplementary Methods). Further dimensionality reduction analysis by t-Distributed Stochastic Neighbor Embedding (t-SNE) (Supplementary Methods) yielded similar findings for each specie (Supplementary Fig. S1A). For the integration of human and mouse cells, we found that samples clustered by disease within species clusters (Supplementary Fig. S1A). Phylogenetic analysis (Supplementary Methods) of the different hematological subtypes and species according to the DNA methylation landscape provided additional evidence of the characteristic epigenetic blueprint of each category (Fig. 1C). The phylogenetic tree reinforced the differences between human and mouse DNA methylomes. Additionally, we found that in humans, hematological cell lines derived from B-cells shared the same node origin in the phylogenetic tree. We next performed a supervised differential methylation analysis (Supplementary Methods) between leukemia, lymphoma and multiple myeloma derived cell lines for human and mouse and we obtained a list of differentially methylated CpG sites (available as Dataset S2). We plotted the methylation β -values of these CpG sites in a heatmap where hematological cell lines were hierarchically clustered and nearly perfectly separated according to each hematological malignancy (Fig. 1D). These differentially methylated CpG sites between the three pathological entities were enriched in open sea regions and promoter regions. Pathway enrichment analyses (Supplementary Methods) unveiled an enrichment in pathways related to phosphorylation regulation and transcription processes (Supplementary Fig. S1B, C). The described clusters in Fig. 1B and D were also observed without data imputation (Supplementary Fig. S2). For both species, normal B and T-cells and myeloid cells clustered apart from their derived malignancies (Supplementary Fig. S3) and the corresponding differential CpG sites are available in Supplementary Dataset S3.

¹Cancer Epigenetics Group, Josep Carreras Leukaemia Research Institute (IJC), Badalona, Barcelona, Catalonia, Spain. ²Centro de Investigacion Biomedica en Red Cancer (CIBERONC), 28029 Madrid, Spain. ³Department of Biosciences, Faculty of Science, Technology and Engineering, University of Vic - Central University of Catalonia (UVic-UCC), Vic, Barcelona, Catalonia, Spain. ⁴Barcelona Supercomputing Center (BSC), Barcelona, Catalonia, Spain. ⁵Department of Hematology, Catalan Institute of Oncology (ICO), Hospital Germans Trias i Pujol, Badalona, Barcelona, Catalonia, Spain. ⁶School of Medicine, Universitat Autònoma de Barcelona, Badalona, Barcelona, Catalonia, Spain. ⁷Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Catalonia, Spain. ⁸Physiological Sciences Department, School of Medicine and Health Sciences, University of Barcelona (UB), Barcelona, Catalonia, Spain. ✉email: mesteller@carrerasresearch.org

Received: 3 July 2024 Revised: 8 November 2024 Accepted: 14 November 2024

Published online: 22 November 2024

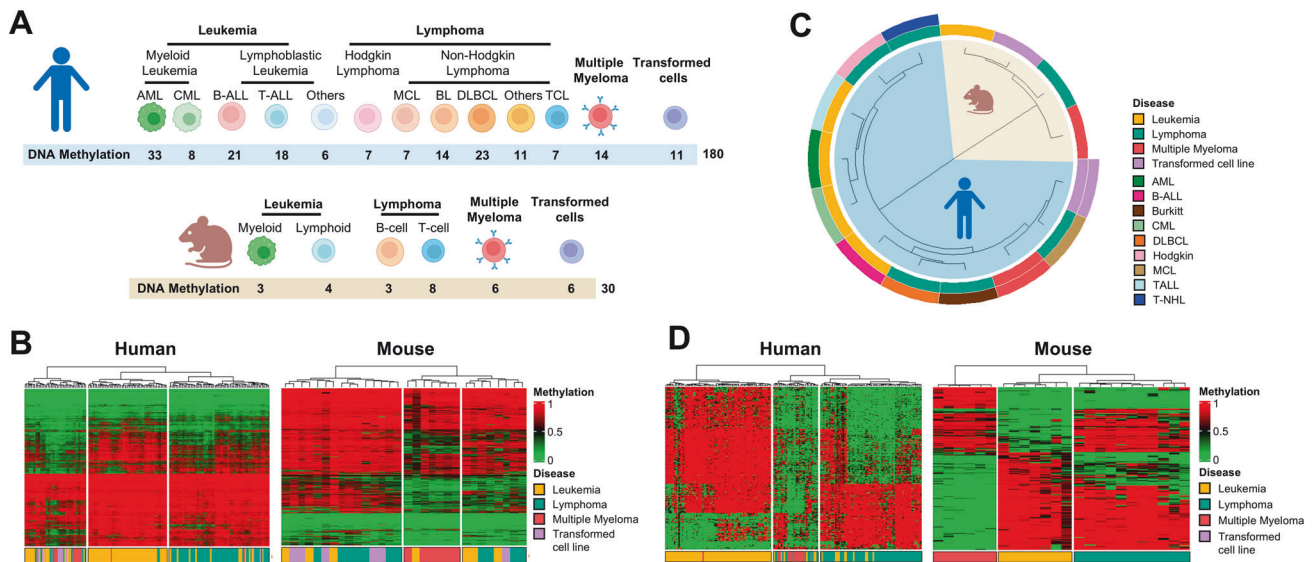


Fig. 1 DNA methylation landscape of human and mouse hematological cell lines. **A** Schematic representation of hierarchical subdivision of human (top) and mouse (bottom) cell line cohorts based on hematological malignancies. **B** Heatmap resulting from the unsupervised analysis of the human (left) and mouse (right) hematological cell lines. Dendrogram clustering was performed using all methylation values, an 1% of randomly selected CpG probes were used for heatmap visualization. **C** Phylogenetic analysis of the different hematological subtypes and species. Each leaf represents the 1% most variable CpG sites within each hematologic subtype and specie. The inner circle colors represent if the leaf corresponds to leukemia, lymphoma, multiple myeloma or transformed cell lines, while the outer circle corresponds to the refined classification for human hematological cell lines. **D** Heatmap resulting from the differential methylation analysis within leukemia, lymphoma and multiple myeloma in human (left) and mouse (right) hematological cell lines. The bottom annotation indicates hematological disease, as described in the heatmap legend. Methylation β -values range from 0 (green) to 1 (red).

Focusing on the human samples, we further examined the granularity of hematological cell lines by performing supervised differential methylation analyses and obtaining lists of CpG methylation sites characteristic of each specific disease in comparison to the others (shown in Supplementary Dataset S4). We used these CpG sites to perform a supervised hierarchical clustering that enabled an effective discrimination of acute myeloid leukemia (AML), chronic myeloid leukemia (CML), B-cell acute lymphoid leukemia (B-ALL), T-cell acute lymphoid leukemia (T-ALL), mantle cell lymphoma (MCL), Burkitt's lymphoma (BL), diffuse large B-cell lymphoma (DLBCL), Hodgkin lymphoma (HL), T-cell lymphoma (TCL) and multiple myeloma (MM) in the heatmap representation (Fig. 2A). The genomic loci and CpG content of the DNA methylation sites that differentiate among the mentioned ten hematological malignancies are shown in Supplementary Fig. 4A. Additional dimensionality reduction analysis using t-SNE (Supplementary Fig. 4B) provided similar results than the supervised hierarchical clustering analysis. The pathway enrichment analyses that were significantly enriched in these ten pathological entities are shown in Supplementary Fig. 4C, mainly involved in T-cell receptor signaling, phosphorylation signaling regulation and transcription regulation. Most important, we translated these in vitro findings of CpG sites that are able to distinguish across different hematological malignancies to the primary setting (Supplementary Methods), using the same DNA methylation microarray platform. Using the above identified CpG sites (Supplementary Dataset S4), the primary cases of AML, B-ALL, T-ALL and DLBCL were clustered almost perfectly in the supervised hierarchical clustering analysis (Fig. 2B) (EGA repository: EGAS50000000627; <https://ega-archive.org/studies/EGAS50000000627>). Interestingly, unsupervised clustering analysis showed that cell lines and primary samples intermingled in many occasions, suggesting DNA methylation resemblance (Supplementary Fig. S5). Those CpG sites with distinct methylation content in vitro vs in vivo are available at Supplementary Dataset S5.

Finally, since expression profiles and IC50 values against hundreds of drugs are available for human hematological cell lines [3, 4, 6], we were able to perform an initial pharmacogenetics exploratory analysis. DNA methylation status of particular genes, such as the case of the DNA repair gene *MGMT* in gliomas, is currently used in precision cancer medicine [11]; and the epigenetic database for hematological malignancies, herein reported, could exhibit translational value. Using the experimental and bioinformatic pipeline described in Supplementary Methods for drugs that are used for hematological malignancies, we were able to identify CpG sites which methylation status and associated expression levels were linked to drug sensitivity (Supplementary Dataset S6). The example of enhanced response to the nucleoside analogs cytarabine, fludarabine and nelarabine [12] and to azacitidine according to the methylation levels sites is illustrated in Fig. 2C and Supplementary Fig. S6, respectively. The DNA methylation patterns of the CpG sites associated with the nucleoside analogs enabled a classification of hematological cell lines as sensitive or resistant through a hierarchical clustering analysis, as it is also shown in the IC50s boxplots in Fig. 2D. For azacitidine, a hypermethylated CpG site within the tumor necrosis factor- α -induced protein 3 gene (*TNFAIP3*) gene was associated with RNA downregulation and enhanced sensitivity (Supplementary Fig. S6). This CpG site was not present in the DNA methylation signature associated with sensitivity to nucleoside analogs, including cytarabine.

Overall, we have demonstrated how human and mouse cell lines derived from hematological malignancies exhibit unique DNA methylation profiles. These epigenetic fingerprints are so characteristic that allowed the obtention of a DNA methylation classifier for ten pathological categories. Most importantly, these data, now freely available to all researchers as a resource, can be further mined for various applications. This is illustrated by our preliminary results, which suggest a potential pharmacogenetic use that could pave the way for translation into clinical studies.

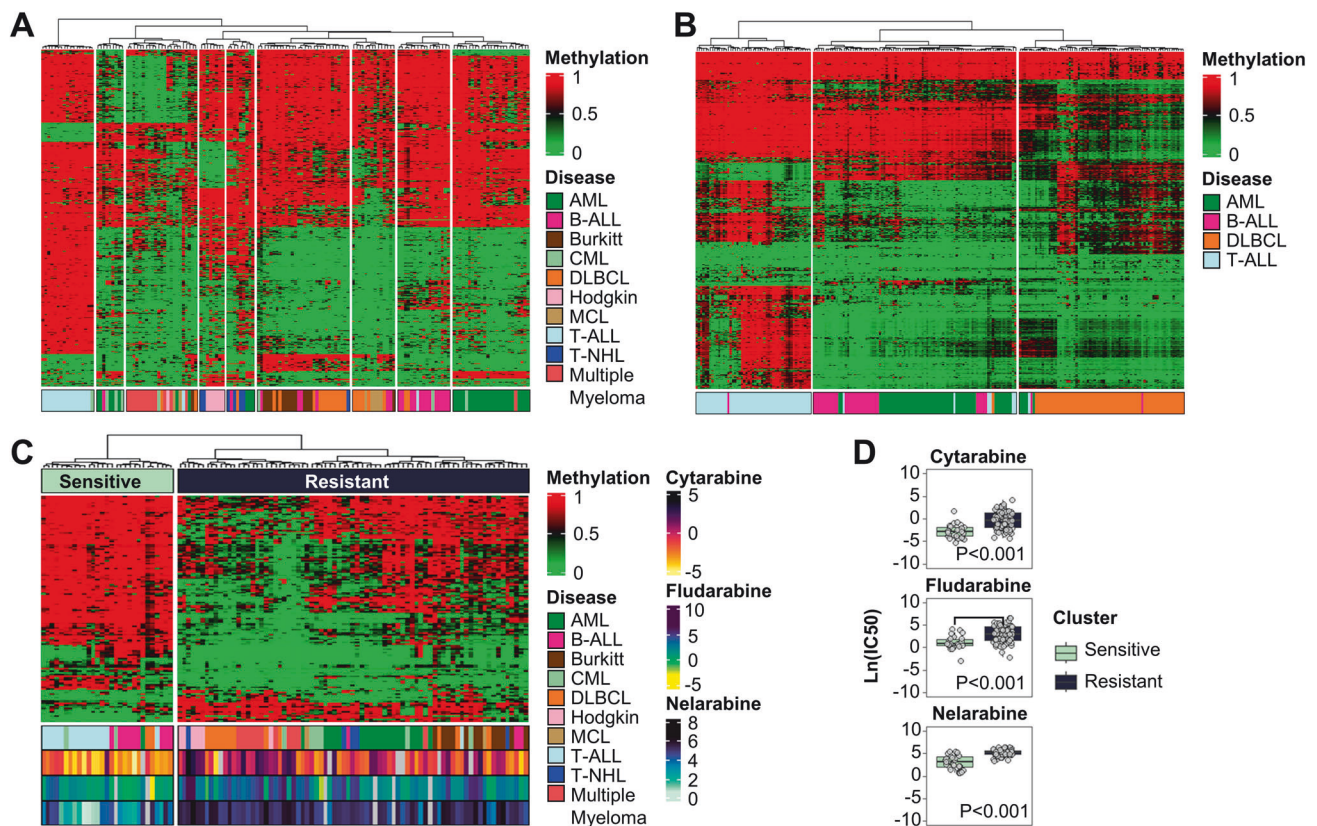


Fig. 2 The DNA methylome of human hematological cell lines can differentiate different hematological malignancies and is associated with drug sensitivity. **A** Heatmap resulting from the CpG sites differentially methylated within hematological diseases in human hematological cell lines. The bottom annotation indicates hematological disease, as described in the heatmap legend. Methylation β -values range from 0 (green) to 1 (red). **B** Heatmap resulting from the CpG sites differentially methylated within hematological diseases in human primary samples. The bottom annotation indicates hematological disease, as described in the heatmap legend. Methylation β -values range from 0 (green) to 1 (red). **C** Heatmap resulting from the common CpG sites associated with cytarabine, fludarabine and nelarabine drug sensitivity. Top annotation indicates whether the cells are in the sensitive cluster or resistant cluster. The bottom annotation indicates hematological disease and cytarabine, fludarabine and nelarabine sensitivity as described in the heatmap legend. Methylation β -values range from 0 (green) to 1 (red). **D** Boxplot representing cytarabine (top), fludarabine (middle) and nelarabine (bottom) IC₅₀s of human hematological cell lines grouped according to whether they clustered in sensitive or resistant cluster. Two-sided Mann–Whitney–Wilcoxon test was performed.

REFERENCES

- Scherer WF, Syverton JT, Gey GO. Studies on the propagation in vitro of poliomyelitis viruses. IV. Viral multiplication in a stable strain of human malignant epithelial cell (strain HeLa) derived from an epidermoid carcinoma of the cervix. *J Exp Med*. 1953;97:695.
- Barretina J, Caponigro G, Stransky N, Venkatesan K, Margolin AA, Kim S, et al. The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature*. 2012;483:603–7.
- Iorio F, Knijnenburg TA, Vis DJ, Bignell GR, Menden MP, Schubert M, et al. A Landscape of Pharmacogenomic Interactions in Cancer. *Cell*. 2016;166:740–54.
- Ghandi M, Huang FW, Jané-Valbuena J, Kryukov GV, Lo CC, McDonald ER 3rd, et al. Next-generation characterization of the Cancer Cell Line Encyclopedia. *Nature*. 2019;569:503–8.
- Bashi AC, Coker EA, Bulusu KC, Jaaks P, Crafter C, Lightfoot H, et al. Large-scale Pan-cancer Cell Line Screening Identifies Actionable and Effective Drug Combinations. *Cancer Discov*. 2024;14:846–65.
- Reinhold WC, Varma S, Sunshine M, Rajapakse V, Luna A, Kohn KW, et al. The NCI-60 Methylome and Its Integration into CellMiner. *Cancer Res*. 2017;77:601–12.
- Sandoval J, Heyn H, Moran S, Serra-Musach J, Pujana MA, Bibikova M, et al. Validation of a DNA methylation microarray for 450,000 CpG sites in the human genome. *Epigenetics*. 2011;6:692–702.
- BLUEPRINT consortium. Quantitative comparison of DNA methylation assays for biomarker development and clinical applications. *Nat Biotechnol*. 2016;34:726–37.
- Moran S, Arribas C, Esteller M. Validation of a DNA methylation microarray for 850,000 CpG sites of the human genome enriched in enhancer sequences. *Epigenomics*. 2016;8:389–99.
- García-Prieto CA, Álvarez-Erriro D, Musulen E, Bueno-Costa A, N Vazquez B, Vaquero A, et al. Validation of a DNA methylation microarray for 285,000 CpG sites in the mouse genome. *Epigenetics*. 2022;17:1677–85.
- Davalos V, Esteller M. Cancer epigenetics in clinical practice. *CA Cancer J Clin*. 2023;73:376–424.
- Tsesmetzis N, Paulin CBJ, Rudd SG, Herold N. Nucleobase and Nucleoside Analogues: Resistance and Re-Sensitisation at the Level of Pharmacokinetics, Pharmacodynamics and Metabolism. *Cancers*. 2018;10:240.

ACKNOWLEDGEMENTS

We want to particularly acknowledge the patients and the Spanish National DNA Bank Carlos III of the University of Salamanca (PT17/0015/0044), integrated in the Spanish National Biobanks Network for their collaboration; the la Fe Biobank (B.0000723) and the IGTP/UJC Biobank. We thank CERCA Programme/Generalitat de Catalunya for institutional support. The Secretariat for Universities and Research of the Ministry of Business and Knowledge of the Government of Catalonia has provided funding to ME (2021 SGR01494). ME has also received funding from the Spanish Ministry of Science and Innovation MCIN/AEI/10.13039/501100011033/ERDF 'A way to make Europe' (PID2021-125282OB-I00), Cellex Foundation (CEL007) and 'la Caixa' Foundation (LCF/PR/HR22/00732). GF is recipient of Ayuda Investigador AECC 2023 (INVES234765FERR), Fundación Científica AECC. ME is an ICREA Research Professor. EG received funding from ISCIII (P119/01828; and P122/01880).

AUTHOR CONTRIBUTIONS

Contribution: AN-C and ME conceived and designed the study; AN-C, CG-P and FS performed molecular and cellular analyses; GF and VD provided hematological cell line classification; VD, EG, JR, and JMR provided primary clinical sample classification; and ME wrote the manuscript with contributions and approval from all authors.

COMPETING INTERESTS

ME declares past grants from Ferrer International and Incyte and personal fees from Quimatrix, outside the submitted work.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41375-024-02478-2>.

Correspondence and requests for materials should be addressed to Manel Esteller.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2024