Haematologica
HAEMATOL/2022/282305
Version 3

Epigenome profiling reveals an aberrant DNA methylation signature in GATA2 deficiency

Oskar Marin-Bejar, Damia Romero-Moya, Javier Rodriguez-Ubreva, Maximiliano Distefano, Francesca Lessi, Paolo Aretini, Alessandro Liquori, Julio Castaño, Emilia Kozyra, Lili Kotmayer, Clara Bueno, José Cervera, José Carlos Rodriguez-Gallego, Josep F Nomdedeu, Laura Murillo-Sanjuán, Cristina Díaz de Heredia, Antonio Pérez-Martínez, Félix López-Cardenas, Carolina Martínez-Laperche, Nieves Dorado-Herrero, Francisco M Marco, Felipe Prósper, Pablo Menendez, David Valcárcel, Esteban Ballestar, Csaba Bödör, Anna Bigas, Albert Catalá, Marcin W Wlodarski, and Alessandra Giorgetti

Disclosures: The authors declare no conflict of interest.

Contributions: OMB, AG, AC and MMW designed the study and wrote the manuscript. OMB, DRM, JR, MD, FL, PA, AL, JC, JP, WK, LK, EB, and AB performed genomic and epigenomic studies and data analysis. JCRG, JFN, CDH, APM, MDC, CML, ND, FM, FP, CB, PM, DV, and AC were involved in the patient care, sample collection, testing and reporting. All authors contributed to the manuscript and provided final approval

Epigenome profiling reveals aberrant DNA methylation signature in GATA2 deficiency

Oskar Marin-Bejar[1#], Damia Romero-Moya[1], Javier Rodriguez-Ubreva[2], Maximiliano Distefano[3], Francesca Lessi[4], Paolo Aretini[4], Alessandro Liquori[5,6], Julio Castaño[7], Emilia Kozyra[8], Lili Kotmayer[9], Clara Bueno[10], José Cervera[5,6,11], José Carlos Rodriguez-Gallego[12,13,14], Josep F Nomdedeu[15], Laura Murillo- Sanjuán[16], Cristina Díaz de Heredia[16], Antonio Pérez-Martinez[17,18,19], Félix López-Cadenas[20,21], Carolina Martínez-Laperche[22], Nieves Dorado-Herrero[22], Francisco M Marco[23], Felipe Prósper[6,24,25] Pablo Menendez[6,10,26,27], David Valcárcel[28], Esteban Ballestar[2,29], Csaba Bödör[9], Anna Bigas[6,30,31], Albert Catalá[3,32], Marcin W Wlodarski[33] and Alessandra Giorgetti[1,4,34#]

1. Regenerative Medicine Program, Bellvitge Institute for Biomedical Research (IDIBELL) and Program for Clinical Translation of Regenerative Medicine in Catalonia (P-CMRC), Barcelona, Spain
2. Epigenetics and Immune Disease Group, Josep Carreras Research Institute (IJC), Barcelona, Spain
3. Department of Hematology and Oncology, Institut de Recerca Sant Joan de Déu, Hospital Sant Joan de Déu, Barcelona, Spain
4. Fondazione Pisana Per la Scienza ONLUS (FPS), San Giuliano Terme, Italy
5. Hematology Research Group, Instituto de Investigación Sanitaria La Fe, Valencia, Spain
6. Centro de Investigación Biomédica en Red de Oncología (CIBERONC), Instituto de Salud Carlos III, Madrid, Spain
7. Advanced and Cell Therapy Services. Banc de Sang i Teixits, Barcelona, Spain
8. Division of Pediatric Hematology and Oncology, Department of Pediatrics and Adolescent Medicine, Medical Center, Faculty of Medicine; Faculty of Biology, University of Freiburg, Freiburg, Germany.
9. HCEMM-SE Molecular Oncohematology Research Group, Department of Pathology and Experimental Cancer Research, Semmelweis University. Budapest, Hungary
10. Josep Carreras Leukaemia Research Institute. Department of Biomedicine. School of Medicine, University of Barcelona, Barcelona, Spain
11. Genetics Unit, Hospital Universitario y Politécnico La Fe, Valencia, Spain
12. Department of Immunology, University Hospital of Gran Canaria Dr. Negrin, Canarian Health System, Las Palmas de Gran Canaria, Spain
13. Department of Clinical Sciences, University Fernando Pessoa Canarias, Las Palmas de Gran Canaria, Spain
14. Department of Medical and Surgical Sciences, School of Medicine, University of Las Palmas de Gran Canaria, Las Palmas de Gran Canaria, Spain
15. Servei d´Hematologia Hospital de la Santa Creu i Sant Pau, Universitat Autònoma de Barcelona, IIB Sant Pau/Josep Carreras Leukaemia Research Institute (IJC), Barcelona, Spain
16. Pediatric Hematology and Oncology Division, Hospital Universitari Vall d'Hebron, Vall d'Hebron Institut de Recerca, Barcelona, Spain
17. Pediatric Department, Universidad Autonoma de Madrid, Madrid, Spain
18. Hospital La Paz Institute for Health Research, Madrid, Spain

19. Pediatric Hemato-Oncology Department, University Hospital La Paz, Madrid, Spain
20. Servicio de Hematología Hospital Clínico Universitario de Salamanca, salamanca, Spain
21. Instituto Biosanitario de Salamanca (IBSAL). Salamanca, Spain
22. Servicio de Hematología, Hospital General Universitario Gregorio Marañón. Instituto de Investigación Sanitaria Gregorio Marañón, Madrid, Spain
23. Immunology Department, Dr. Balmis General University Hospital; Institute for Health and Biomedical Research (ISABIAL), Alicante, Spain
24. Area de Hemato-Oncología, CIMA Universidad de Navarra, Instituto de Investigación Sanitaria de Navarra (IDISNA), Pamplona, Spain
25. Servicio de Hematologia, CCUN, Clínica Universidad de Navarra, Universidad de Navarra, Pamplona, Spain
26. Red Española de Terapias Avanzadas (TERAV) - Instituto de Salud Carlos III, Madrid, Spain
27. Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Spain
28. Servei d'Hematologia, Vall d'Hebron Hospital Universitari; Experimental Hematology, Vall d'Hebron Institute of Oncology (VHIO); Departament de Medicina, Universitat Autònoma de Barcelona, Barcelona, Spain
29. Epigenetics in Inflammatory and Metabolic Diseases Laboratory, Health Science Center (HSC), East China Normal University (ECNU), Shanghai, China
30. Programa de Investigación en Cáncer, IMIM (Hospital del Mar Medical Research Institute), Barcelona, Spain
31. Josep Carreras Leukaemia Research Institute (IJC), Barcelona, Spain
32. Centro de Investigación Biomédica en Red de Enfermedades Raras (CIBERER), Instituto de Salud Carlos III, Madrid, Spain
33. Department of Hematology, St. Jude Children's Research Hospital, Memphis, TN, USA
34. Department of Pathology and Experimental Therapeutics, Faculty of Medicine and Health Sciences, Barcelona University, Barcelona, Spain

Corresponding author:
Alessandra Giorgetti
Address: Bellvitge Institute for Biomedical Research (IDIBELL)
Hospital Duran i Reynals-3rd floor
Gran Via de l'Hospitalet, 199
08908 Hospitalet de Llobregat - Barcelona - Spain
agiorgetti@idibell.cat

Co-corresponding author: Oskar Marin Bejar
Address: Bellvitge Institute for Biomedical Research (IDIBELL)
Hospital Duran i Reynals-3rd floor
Gran Via de l'Hospitalet, 199
08908 Hospitalet de Llobregat - Barcelona - Spain
omarin@idibell.cat

Contributions

OM-B, AG, AC and MWW designed the study and wrote the manuscript. OM-B, DRM, JR, MD, FL, PA, AL, JC, JP, WK, LK, EB, and AB performed genomic and epigenomic studies and data analysis. JCRG, JFN, CDH, APM, MDC, CML, ND, FM, FP, CB, PM, DV, and AC were involved in the patient care, sample collection, testing and reporting. All authors contributed to the manuscript and provided final approval

Disclosure

The authors declare no conflict of interest.

Acknowledgment

We thank Francesca di Giorgio, Loris Mularoni, Chiara Mazzanti, Joan Pera, Dolly Viviana Fiallo-Suárez, Adela Escudero-Lopez, María Díez-Campelo, Teresa González, Montserrat Arnan, Francesc Solé, Amaia Vilas Zornoza, Montse Rovira (Hospital Clinic, Barcelona), Paola Romecin and Laura Palomo, for technical support.

Funding

This work was supported by ERA PerMed GATA2-HuMo Funding Mechanism (Spain: Acció instrumental de SLT011/18/00006 of the Department of Health of the Government of Catalonia, to AG and AB; Hungary: ED-18- 1-2019-001 grant from the National Research, Development and Innovation Office to CB; and Germany: German Federal Ministry of Education and Research (BMBF) 2018-123/01KU1904 to MWW), ÚNKP-21-2-I-SE-21 and Hungarian National Academy of Scientist Education grant to KL, TKP2021-NVA-15, TKP2021-EGA-24 and EU's Horizon 2020 research and innovation program under grant agreement no. 739593 and Elixir Hungary to CB, the Spanish Ministry of Economy, Industry, CERCA/Generalitat de Catalunya and Fundació Josep Carreras-Obra Social la Caixa and the Deutsche Josep Carreras Leukämie-Stiftung (DJCLS15R/2021) to PM, Asociación Española contra el cancer (AECC, PRYGN211192BUEN) and Health Institute Carlos III (PI20/00822) to CB, Competitiveness (MINECO PID2020-15591RB-100), La Marató de TV3 (202001-32), FPS Grant 2018 by Fondazione Pisana per la Scienza ONLUS and CERCA Programme/ Generalitat de Catalunya for institutional support to AG, and BMBF MyPred 01GM1911A to MWW. OM-B Is supported by 101029927-scGATA2track (H2020-MSCA-IF-2020) and KOG-202109-01162, AL is the recipient of the APOSTD2021/212 fellowship from the Generalitat Valenciana.

Data sharing statement

The sequencing data has been deposited to EGA European Genome-Phenome Archive, genomic WES data, PRJEB58433 and to GEO, epigenome data, GSE221745.

GATA2 deficiency is a complex multi-system disorder with high risk of developing myelodysplastic syndromes (MDS) and acute myeloid leukemia (AML) with a nearly complete lifetime penetrance[1, 2]. GATA2 carriers show a highly variable expressivity, with some individuals developing early-onset MDS, while others, remaining asymptomatic throughout life. Although no prognostic biomarkers exist, it is likely that both cooperating genetic and epigenetic drivers shape the course of the disease[3]. Despite advances in the identification of recurrent somatic mutations in a set of leukemia driver genes (i.e. *STAG2, SETBP1, ASXL1* and *ETV6*), there are major gaps in understanding the molecular mechanisms associated with leukemic progression in GATA2 carriers[4]. Moreover, DNA methylation alterations contribute to the initiation and expansion of leukemic clones and aberrant hypermethylation occurs in adult patients with MDS and AML[5, 6]. However, to date, a genome-wide DNA methylome analysis in GATA2 patients has not been performed.

For this study, 20 clinically annotated GATA2 carriers from seven Spanish hospitals were enrolled, (Table 1 and *Online Supplementary Figure S1)*. Median age at the diagnosis was 36 (6–75) years. The primary initial manifestation was MDS (n=12, 55%), followed by immunodeficiency (n=3, 15%), and AML (n=2, 10%). On cytogenetics, trisomy 8 was detected in 3 patients, complex karyotype in 2 patients, while 11 patients had normal karyotype (*Online Supplementary Figure S1A-F*). Based on DNA availability, somatic mutation profiling was performed in total peripheral blood (PB) or bone marrow (BM) of 17 GATA2 carriers. Somatic mutations in myeloid malignancy genes were identified in 71% (12/17) tested patients (*Online Supplementary Figure S1F)*. This analysis confirmed the heterogeneity of acquired somatic mutations in GATA2 deficiency, with *STAG2*, *ASXL1* and *SETBP1* as recurrently affected genes. We obtained global DNA methylation profiles of patients who underwent complete genetic characterization by using the Infinium Human Methylation EPIC 850□K platform (Illumina). We profiled DNA of 8 BM (P1, P3, P5, P6, P10, P11, P12, P13) and 8 PB samples (P1, P4, P7, P8, P9, P13, P16 and P17) and compared with a cohort of 12 (5 PB and 7 BM) age-matched healthy donor (HD) controls.

High-dimensional data visualization showed that the majority of GATA2 patients cluster tightly together and separately from HD. Asymptomatic carriers P1, P16 and P17 (at age 6, 40 and 75 years old, respectively) belonging to the same family, were encompassed to HD group (Figure 1A). Additionally, we compared HD with the asymptomatic GATA2 carriers alone, and we still observed intermixed samples (data not shown). Next, the DNA methylation changes were calculated between GATA2 group and HD pairwise, revealing a DNA methylation pattern specific to GATA2 carriers. In detail, 2834 differentially methylated positions (DMPs) were identified in GATA2 BM samples and 1406 DMPs in PB samples (*Online Supplementary Figure S2A*). A descriptive analysis of the DMP distribution was performed using as a reference the probe distribution of the Infinium MethylEPIC array from distal to proximal CpG island regions (open sea, CpG shelf, CpG shore and CpG island) (*Online Supplementary Figure 2B*). Although previous studies showed that promoter-proximal methylation is negatively correlated with active gene expression[7]; our analysis revealed that the majority of DMPs are promoter-distant in both, BM (Open Sea: 60% hypomethylated DMPs and 72.5% hypermethylated DMPs) and PB samples (Open Sea: 52% hypomethylated and 82.3% hypermethylated) (Figure 1B and *Online Supplementary Figure 2C*).

Additionally, the DMP distribution using the neighboring gene as reference showed an enrichment in intergenic and intronic regions (Figure 1C, *Online Supplementary Figure S2D-E*). This observation was corroborated by a correlation analysis comparing the DMP distribution of GATA2 patients with the reference array (*Online Supplementary Figure S2F*). Overall, we observed that DNA methylation changes are enriched in gene-distant and intronic regions in GATA2 patients. Whether this discrepancy is a consequence of GATA2 deficiency, or arising solely from the MDS evolution, needs to be further investigated. Previous studies showed that endogenous GATA2 preferentially occupies sites distant to promoters in hematopoietic stem cells[8], hence the loss of DNA binding capacity of GATA2 mutant protein might result in aberrant DNA methylation of these loci.

Interestingly, unsupervised analysis of hypermethylated DMPs highlighted the presence of a hypermethylated DMP subcluster across all the GATA2 BM samples (hereinafter subcluster A), including P1, the asymptomatic GATA2 carrier (Figure 1D). The matching PB of P1 revealed a subcluster of hypermethylated DMPs (hereinafter subcluster B) as affected *GATA2* patients (Figure 1E). Importantly, the 2-year longitudinal follow-up of P1 showed the evolution to MDS with multilineage dysplasia (MDS-MLD) and monosomy 7 without secondary mutations. In contrast, the two asymptomatic GATA2 carriers P16 and P17 (father and grandfather of P1) had a DNA methylation profile comparable with the HD group (Figure 1E). This observation suggests the presence of a likely early aberrant DNA methylation at specific loci in GATA2 carriers that might have a potential use in early detection of patients at risk for impending myeloid transformation.

After the gene annotation of the 131 DMPs of the subcluster A (*Online Supplementary Figure S2G)*, 118 genes were associated to hypermethylated DMPs (*Online Supplementary Figure S2H)*, including *MECOM*, which epigenetic regulation has been reported in AML[9]. The top candidate was *PROMININ1* (PROM1/CD133) with four hypermethylated DMPs upstream of its promoter region (*Online Supplementary Figure S2I*). Interestingly, the aberrant methylation status of *PROM1* promoter has already been described in various cancers including AML[10].
Moreover, 42 out of 131 hypermethylated DMP subclusters are classified as gene regulatory elements; and most of them are enriched for H3K27ac, a chromatin mark associated with enhancer activity (Figure 2A).

Because tissue-specific DNA methylation patterns might alter the methylation results, we overlapped the 118 BM-hypermethylated genes with the 1060 PB-hypermethylated genes. We found 51 commonly hypermethylated common genes, implying that PB samples, at least partially reflect the dysregulated pattern observed in the BM (Figure 2B). Additionally, the 205 PB-hypermethylated subcluster B genes were crossed with 118 BM-hypermethylated subcluster A genes, observing an overlapped of 30 genes, indicating that the aberrant DNA methylation profile is detectable in both BM and PB samples (*Online Supplementary Figure S2H*). Aberrant epigenetic changes have been associated with alterations of transcription factors (TFs) genomic binding capacities[11,12]. Therefore, we assessed whether the hypermethylated and hypomethylated DMPs identified in PB and BM samples were enriched in specific TF DNA binding motives using Hypergeometric Optimization of

Motif EnRichment (HOMER). This analysis revealed in the hypermethylated DMPs a significant enrichment in TFs motives of the ETS family, such as ETV2, ETV6, ELF5 and PU.1 among others (Figure 2C-D), which are known to play a role in MDS[13]. The hypomethylated DMPs showed an enrichment in TF of the bZIP family preferentially (*Online Supplementary Figure S3A-B*). To evaluate whether the GATA2 binding sites are linked to DNA methylation in GATA2 deficiency, we integrated the hypermethylated genes in both PB and BM samples with a GATA2 ChIPseq dataset from our laboratory (GSE107639)[14]. This analysis revealed the presence of 82 out of 494 hypermethylated genes that are also GATA2-targets (Figure 2E). On the contrary, the intersection of GATA2-regulated genes with the hypomethylated genes in BM and PB did not show any relevant gene enrichment (*Online Supplementary Figure S3C*). Additionally, the common hypermethylated genes of PB and BM were crossed with K562 myeloid leukemia GATA2 ChIPseq dataset (GSE18868)[8], observing that 204 out of 494 hypermethylated genes are GATA2 targets and 51 out of 82 genes are in common between Castaño *et al*[14] and Fujiwara *et al*[8] datasets *(Online Supplementary Figure S3D-E)*. Gene ontology analysis of the 82 hypermethylated GATA2 target genes showed an enrichment in transcriptional regulation and cell differentiation. *In-silico* analysis of the upstream regulators inferred that ETV6, TCF12, MGA, and SOX5 are cooperative-TFs of GATA2 gene regulatory network *(Online Supplementary Figure 3F)*. Together, our data suggest that GATA2 deficiency is associated with aberrant DNA methylation in GATA2 target genes.

Finally, we compared our BM hypermethylation data with publicly available methylation profiles of 184 pediatric AML patients, TARGET 2018[15]. Interestingly, 50% (4/8) of our GATA2 patients clustered together with AML samples (which had known GATA2 mutation negative status), showing a similar methylation pattern with AML samples (*Online Supplementary Figure S3G*). This points to the possibility that some aberrant methylation signatures observed in our GATA2 patients might be directly linked to the AML transformation and thus arises independently of the underlying GATA2 germline mutation. Future genome-wide association studies are warranted to address this question in depth.

In conclusion, we identified an aberrant DNA hypermethylated signature in GATA2 deficiency. Specifically, we described the presence of a subset of aberrant hypermethylated set of genes present in GATA2 carriers at early (and not yet symptomatic) disease stage, which could be potentially used as predictors of disease progression. In this context, the implementation of customized methylation-specific assays might be instrumental to validate our findings in larger cohorts of patients and to test its clinical prognostic utility. Finally, a collaborative effort will be essential to increase the number patients with this rare yet high-risk MDS/AML predisposition syndrome, allowing for comprehensive genetic and epigenetic analyses to understand the impact of the secondary hits and/or aberrant DNA methylation on the disease progression.

REFERENCES

1. Kotmayer L, Romero-Moya D, Marin-Bejar O, et al. GATA2 deficiency and MDS/AML: Experimental strategies for disease modelling and future therapeutic prospects. 2022.
2. Zhang Y, Wu J, Qin T, et al. Comparison of the revised 4th (2016) and 5th (2022) editions of the World Health Organization classification of myelodysplastic neoplasms. Leukemia. 2022;36(12):2875-2882.
3. Al Seraihi AF, Rio-Machin A, Tawana K, et al. GATA2 monoallelic expression underlies reduced penetrance in inherited GATA2-mutated MDS/AML. Leukemia. 2018;32(11):2502-2507.
4. West RR, Calvo KR, Embree LJ, et al. ASXL1 and STAG2 are common mutations in GATA2 deficiency patients with bone marrow disease and myelodysplastic syndrome. Blood Adv. 2022;6(3):793-807.
5. Figueroa ME, Abdel-Wahab O, Lu C, et al. Leukemic IDH1 and IDH2 mutations result in a hypermethylation phenotype, disrupt TET2 function, and impair hematopoietic differentiation. Cancer Cell. 2010;18(6):553-567.
6. Figueroa ME, Skrabanek L, Li Y, et al. MDS and secondary AML display unique patterns and abundance of aberrant DNA methylation. Blood. 2009;114(16):3448-3458.
7. Futscher BW, Oshiro MM, Wozniak RJ, et al. Role for DNA methylation in the control of cell type specific maspin expression. Nat Genet. 2002;31(2):175-179.
8. Fujiwara T, O'Geen H, Keles S, et al. Discovering hematopoietic mechanisms through genome-wide analysis of GATA factor chromatin occupancy. Mol Cell. 2009;36(4):667-681.
9. Groschel S, Sanders MA, Hoogenboezem R, et al. A single oncogenic enhancer rearrangement causes concomitant EVI1 and GATA2 deregulation in leukemia. Cell. 2014;157(2):369-381.
10. Gopisetty G, Xu J, Sampath D, Colman H, Puduvalli VK. Epigenetic regulation of CD133/PROM1 expression in glioma stem cells by Sp1/myc and promoter methylation. Oncogene. 2013;32(26):3119-3129.
11. Garcia-Gomez A, Li T, de la Calle-Fabregat C, et al. Targeting aberrant DNA methylation in mesenchymal stromal cells as a treatment for myeloma bone disease. Nat Commun. 2021;12(1):421.
12. Blattler A, Farnham PJ. Cross-talk between site-specific transcription factors and DNA methylation states. J Biol Chem. 2013;288(48):34287-34294.
13. Zhao J, Li D, Seo J, Allen AS, Gordan R. Quantifying the Impact of Non-coding Variants on Transcription Factor-DNA Binding. Res Comput Mol Biol. 2017;10229(336-352.
14. Castano J, Aranda S, Bueno C, et al. GATA2 Promotes Hematopoietic Development and Represses Cardiac Differentiation of Human Mesoderm. Stem Cell Reports. 2019;13(3):515-529.
15. Bolouri H, Farrar JE, Triche T, Jr., et al. The molecular landscape of pediatric acute myeloid leukemia reveals recurrent structural alterations and age-specific mutational interactions. Nat Med. 2018;24(1):103-112.

Table 1

ID	GATA2 Mutations		Age	Sex	Clinical manifestation*	Karyotype category	Treatment	Status	Genomic analysis	PB Monocytes (10^3 cells/uL)	PB NK (cells/uL)	PB B cells (cells/uL)	BM Blasts (%)
	Mutation	Mutation Type											
P1	p.M388T	Missense	6	M	Asymptomatic	NK	None	Alive	WES	0,9	200	420	0
P2	p.R362*	Nonsense	18	M	MDS	NK	None	Alive	WES	0,01	NA	93	0
P3	p.K378*	Nonsense	13	F	ID	Complex/+8	HSCT	Alive	WES	NA	NA	NA	0
P4	p.T354M	Missense	37	F	MDS	NA	None	Alive	WES	0,16	7	18	NA
P5	p.G273Dfs*53	Frameshift	51	F	MDS	NK	None	Alive	WES	0,01	NA	NA	3
P6	p.S261T	Missense	75	F	AML	+8	None	Dead	WES	0,06	NA	NA	29
P7	p.R396L	Missense	30	M	MDS	NA	None	Dead	WES	0,29	0	97	NA
P8	p.R396L	Missense	15	F	ID	NA	None	Dead	WES	0,1	16	35	0
P9	p.G346Sfs*40	Frameshift	24	M	MDS	NA	None	Dead	WES	0,198	34	0	NA
P10	p.L305V	Missense	59	F	MDS	NK	HSCT	Alive	WES	0	23	81	1
P11	p.R396Q	Missense	45	M	MDS	Complex/-7q	HSCT	Alive	WES	0	0	30	1
P12	p.R396W	Missense	25	M	MDS	+8	HSCT	Dead	WES	0	1	6	2
P13	p.T358I p.G149R	Missense Missense	55	M	MDS	NK	HSCT	Alive	WES	0,72	8	14	11
P14	p.L386Hfs*2	Frameshift	29	F	MDS	NK	HSCT	Alive	Targeted seq.	0	40	0	1
P15	p.G346insSAA	Insertion	34	F	MDS	NK	HSCT	Alive	Targeted seq.	0	NA	NA	3
P16	p.M388T	Missense	40	M	Deafness	NK	None	Alive	WES	NA	NA	NA	0
P17	p.M388T	Missense	75	M	Asymptomatic	NK	None	Alive	WES	NA	NA	NA	0
P18	p.M388T	Missense	39	F	AML	NK	HSCT	Alive	NA	0	NA	NA	0
P19	p.T117=	Synonymous	49	M	Asymptomatic	NK	None	Alive	NA	0,6	NA	NA	0
P20	p.R396L	Missense	14	M	MDS	NA	NA	Dead	NA	0,07	184	12	NA

Table 1. Patient clinical characteristics and genetic landscape.

Patients were diagnosed in the following hospitals: University Hospital of Gran Canaria Dr. Negrin, Hospital de la Santa Creu i Sant Pau, Hospital Universitari Vall d'Hebron, Hospital La Paz, Hospital Clínico Universitario de Salamanca, Hospital General Universitario Gregorio Marañón and Dr. Balmis General University Hospital. M, male; F, female; NK, normal karyotype; NA, no available; MDS, myeloid dysplastic syndrome; ID, immunodeficiency; AML, Acute Myeloid Leukemia; WES, whole exome sequencing; * at sample collection.

Figure legends:

Figure 1. Unsupervised hierarchical clustering and the heat map visualization of differentially methylated CpG sites of GATA2 patients.

A) T-distributed stochastic neighbor embedding (t-SNE) showing the distribution of GATA2 patient (P) data (in red) and healthy donors (HD) (in blue), based on DNA methylation profile. Bone marrow (BM) samples are represented with circles and peripheral blood (PB) samples are represented with triangles.

B) Differentially methylated probes (DMPs) distribution of BM samples, hypermethylated (above) and hypomethylated (below). CpG island distance, island (eggplant), shore (lollipop), shelf (mauve) and open sea (fandango).

C) DMPs distribution of BM samples, hypermethylated (above) and hypomethylated (below). Promoter (dark blue), intergenic (orange), intronic (grey), exonic (yellow), 5' UTR (light blue) and 3' UTR (green).

D) Heatmap of DMPs BM vs HD samples. The DMPs in common among all GATA2 patients (subcluster A) are squared in green. Scale Beta values from -3 (blue/hypomethylated) to +3 (red/hypermethylated).

E) Heatmap of DMPs of PB vs HD samples. The hypermethylated DMPs in P1 sample (subcluster B), which are in common with affected GATA2 patients, are squared in fuchsia. Scale Beta values from -3 (blue/hypomethylated) to +3 (red/hypermethylated).

Figure 2. Regulatory element analysis of hypermethylated genomic regions identify in GATA-mutant patients.

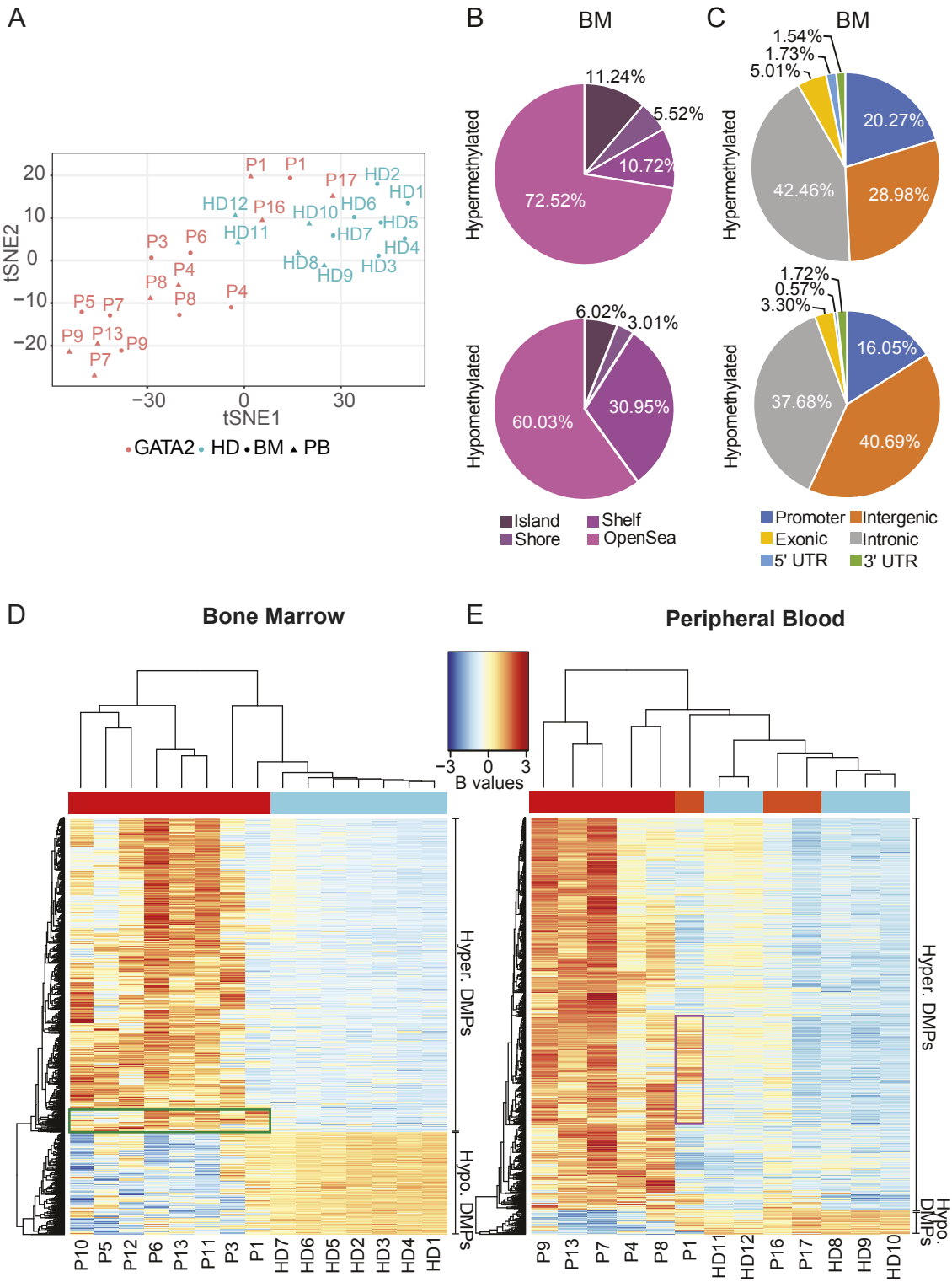
A) Ranked representation of the regulatory function of the hypermethylated position based on GeneHancer (blue) score and H3K27ac enrichment of those genomic positions (green).

B) Venn diagram of the total hypermethylated genes in peripheral blood (PB) (n=1060) versus subcluster of hypermethylated genes in bone marrow (BM) (n=118), the genes of the intersection are 51 genes. P-value=1.572541e-14 (hypergeometric distribution test).

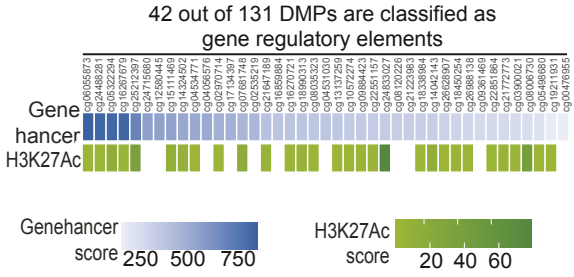
C) Hypergeometric Optimization of Motif EnRichment (HOMER) analysis using hypermethylated differentially methylated probes in BM samples. Enriched motifs found are predominantly from the ETS family including ETVs and PU.1.

D) HOMER analysis using hypermethylated DMPs in peripheral blood (PB) samples. Enriched motifs found are predominantly from the ETS family and IRF family.

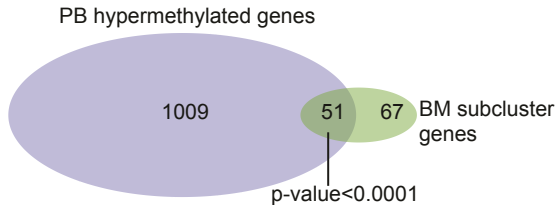
E) Top row: Venn diagram of the neighbouring gene of hypermethylated DMPs in the BM samples (n=1631) compared to the neighbouring gene of hypermethylated DMPs in the PB samples (n=1095), intersection 494 genes. Below row: GATA2 regulated genes (n=2301) obtained from the intersection of two GATA2 ChIPseq data GSE107639[15]. The crossed of top intersection (n=494) versus GATA2-regulated genes (n=2301), the intersection gives 82 GATA2 hypermethylated targets both in PB and BM.


A
tSNE2
tSNE1
GATA2 HD BM PB
B
BM
Hypermethylated
11.24%
5.52%
10.72%
72.52%
Hypomethylated
6.02%
3.01%
30.95%
60.03%
Island
Shore
Shelf
OpenSea
C
BM
Hypermethylated
1.54%
1.73%
5.01%
20.27%
28.98%
42.46%
Hypomethylated
1.72%
0.57%
3.30%
16.05%
40.69%
37.68%
Promoter
Intergenic
Exonic
Intronic
5' UTR
3' UTR
D
Bone Marrow
B values
Hyper. DMPs
Hypo. DMPs
P10 P5 P12 P6 P13 P11 P3 P1 HD7 HD6 HD5 HD2 HD3 HD4 HD1
E
Peripheral Blood
Hyper. DMPs
Hypo. DMPs
P9 P13 P7 P4 P8 P1 HD11 HD12 P16 P17 HD8 HD9 HD10

A

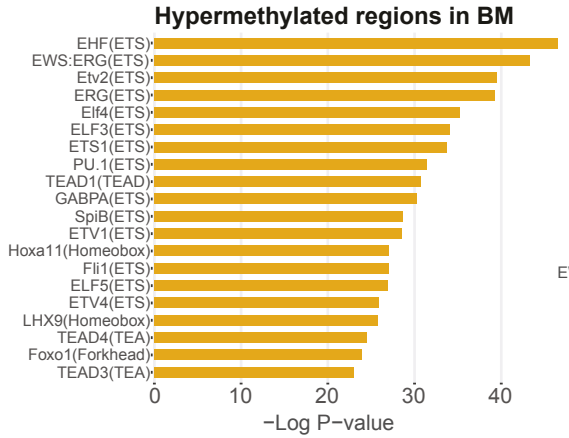


B

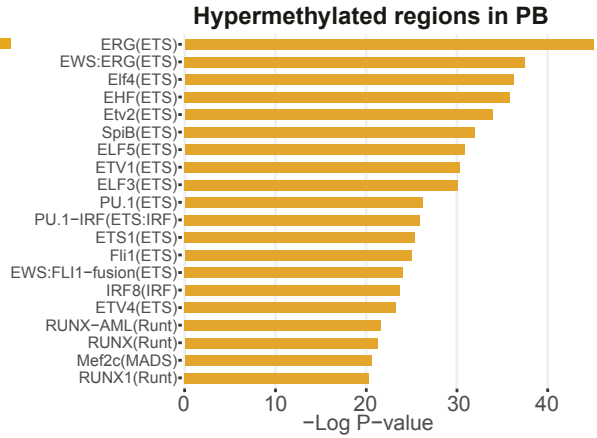


MIR4272	MIR4675	DCDC1	ASB13	FREM1	LINC02271	ETV4	PIMREG	PRICKLE2
ZEB2	ADGRG7	LOX	TTYH3	PDCH9	MAS1L	MIR7850	NRP2	C8ORF34
HPN-AS1	ADAMTS6	MIR1203	SCAF8	DNAH9	TOP1P2	PIK3C2G	PLCB1-IT1	LOC100131257
CALCRL	ABHD15	CNTN4	PPM1L	DIP2B	HTR2A	SNORD151	MGC34796	NRTN
ARHGAP23	BCKDHA	CAMTA1	CD200R1	CDH7	EFL1P1	ZFPM2	C4orf46	LOC105374205
PRDM10	PTN	NXPH1	BNIP3	SPP2	LINC01412			

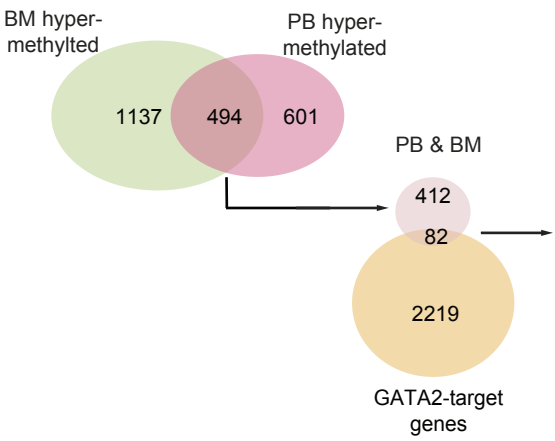
C



D



E



GATA2 targets hypermethylated in PB and BM

SPRY4-AS1	ARHGAP26-IT1	LINC00880	BTG3	KANK1	LIMCH1
FAM169B	TCF4	ST3GAL1	COL23A1	WNT5A	LRRC8B
PRKCE	RAD52	MTMR2	ETV6	TMCC3	FBN2
SORL1	MN1	KDR	PALLD	SORBS1	GBX2
NRP2	LINC00910	MCF2L	MIR6073	CAMTA1	UBE2J1
AHNAK	SLC20A2	AFAP1L2	LINC01798	CXCR4	CCNG2
SOX5	TIPARP	RABGAP1L	FRMD4A	TYW5	LBP
LNPEP	PRDM8	IRAK2	LINC01623	MGA	CYP2C18
LOC101929237	TCF12	CACNB2	CCNG1	LOC102724152	HAS3
ARID1A	EHF	SUMO1P1	C1QTNF7	JARID2	SLC9A7P1
SALL1	SNORD168	SIPA1L1	CUX1	CORO2B	DNAH9
ZNF521	PCDH9	ST6GAL1	TET2	NID2	ARHGAP23
SOCS2	FOXN3	DCC	OXR1	TUBB	
PTP4A2	PCAT1	LOC105377621	DIP2B	ARHGEF3	