Genome-wide regulation of CpG methylation by ecCEBPα in acute myeloid leukemia [version 2; peer review: 2 approved]

Adewale J. Ogunleye¹, Ekaterina Romanova², Yulia A. Medvedeva¹,²

¹Department of Biological and Medical Physics, Moscow Institute of Physics and Technology, Moscow, Russian Federation
²Research Center of Biotechnology, Institute of Bioengineering, Russian Academy of Sciences, Moscow, Russian Federation

**Abstract**

**Background:** Acute myeloid leukemia (AML) is a hematopoietic malignancy characterized by genetic and epigenetic aberrations that alter the differentiation capacity of myeloid progenitor cells. The transcription factor CEBPa is frequently mutated in AML patients leading to an increase in DNA methylation in many genomic locations. Previously, it has been shown that ecCEBPα (extra coding CEBPa) - a lncRNA transcribed in the same direction as CEBPa gene - regulates DNA methylation of CEBPa promoter in cis. Here, we hypothesize that ecCEBPα could participate in the regulation of DNA methylation in trans.

**Method:** First, we retrieved the methylation profile of AML patients with mutated CEBPa locus from The Cancer Genome Atlas (TCGA). We then predicted the ecCEBPα secondary structure in order to check the potential of ecCEBPα to form triplexes around CpG loci and checked if triplex formation influenced CpG methylation, genome-wide.

**Results:** Using DNA methylation profiles of AML patients with a mutated CEBPa locus, we show that ecCEBPα could interact with DNA by forming DNA:RNA triple helices and protect regions near its binding sites from global DNA methylation. Further analysis revealed that triplex-forming oligonucleotides in ecCEBPα are structurally unpaired supporting the DNA-binding potential of these regions. ecCEBPα triplexes supported with the RNA-chromatin co-localization data are located in the promoters of leukemia-linked transcriptional factors such as MLF2.

**Discussion:** Overall, these results suggest a novel regulatory mechanism for ecCEBPα as a genome-wide epigenetic modulator through triple-helix formation which may provide a foundation for sequence-specific engineering of RNA for regulating methylation of specific genes.
Keywords
Acute myeloid leukemia, Triplex, DNA methylation, long non coding RNA, extra-coding CEBPα

Corresponding author: Adewale J. Ogunleye (ogunleie.ad@phystech.edu)

Author roles: Ogunleye AJ: Formal Analysis, Investigation, Methodology, Validation, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing; Romanova E: Conceptualization, Data Curation, Supervision; Medvedeva YA: Conceptualization, Funding Acquisition, Methodology, Resources, Software, Supervision, Validation, Visualization, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

Grant information: The work has been partially supported by Russian Science Foundation (grant 18-14-00240) to YAM and the Ministry of the Sciences and Higher Education of the Russian Federation. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Copyright: © 2021 Ogunleye AJ et al. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Ogunluye Aj, Romanova E and Medvedeva YA. Genome-wide regulation of CpG methylation by ecCEBPα in acute myeloid leukemia [version 2; peer review: 2 approved] F1000Research 2021, 10:204 https://doi.org/10.12688/f1000research.28146.2

First published: 11 Mar 2021, 10:204 https://doi.org/10.12688/f1000research.28146.1
Introduction

Acute myeloid leukemia (AML) is a malignant tumor characterized by the proliferation of undifferentiated myeloblasts. It is the most prevalent form of leukemia in older adults (>60 years) with an annual mortality rate of 50% and a 5-year survival rate of 24%. With the combined effects of the global increase in average life expectancy and AML drug inefficiency, the number of patients is expected to significantly increase in the coming years.

The current understanding of the molecular interplay in AML has been defined under two distinct categories; (i) genetic abnormalities and (ii) non-random chromosomal rearrangements. Cases of AML with chromosomal rearrangements as t(15;17) [PML-RARA], t(9;22) [BCR-ABL], inv(16) [CBFB-MYH11], t(8;21) [RUNX1-ETO] are called cytogenetically abnormal (CA-AML), while cases with genetic abnormalities (including frequent mutations in DNMT3A, NPM1, CEBPa, IDH1/2, TET2, FLT3-ITD) are called cytogenetically normal (CN-AML). The former accounts for 50–55%, while the latter accounts for 45–50% of diagnosed AML cases.

mutations and chromosomal alterations are crucial for initiating AML, they are not sufficient to explain AML progression, heterogeneity, and relapse.

Recently, studies have identified the role of non-coding RNAs, especially long non-coding RNAs (LncRNA) in the initiation and progression of cancers. LncRNAs are emerging functional transcriptional products of at least 200 nucleotides lacking an open reading frame. Although they account for a large proportion of transcriptional products in mammals (about 58,000 loci), only a small number of LncRNAs have been well characterized. Although LncRNAs are mostly not conserved evolutionarily, they are heavily regulated suggesting their functional role. LncRNAs may function either as signal transducers, protein guides, or molecular scaffolds to regulate transcriptional and epigenetic events. Some LncRNAs perform these functions in cis- by modulating transcription of nearby genes (Dum) or act in trans- by modulating genes at multiple distant loci (MALAT1), while some can do both (HOTAIR).

Recent studies have identified LncRNAs for their remarkable role in regulating major epigenetic processes such as DNA methylation and chromatin remodeling. DNA methylation in mammals is coordinated by one of the three DNA methyltransferases (DNMT); DNMT1, DNMT3A, and DNMT3B. LncRNAs have been identified in recent studies as important agents that can modulate DNA methylation, either by activating or repressing DNMTs. For example, the LncRNA Dum was discovered to repress a nearby gene Dppa2 by recruiting multiple DNMTs leading to methylation of a promoter region, thus promoting myoblast differentiation. Conversely, the LncRNA H19 represses the activity of DNMT3B by interacting with SAHH which hydrolyzes SAH, a step required for DNMT3B activation. LncRNAs are emerging functional transcriptional products of at least 200 nucleotides lacking an open reading frame. Although they account for a large proportion of transcriptional products in mammals (about 58,000 loci), only a small number of LncRNAs have been well characterized. Although LncRNAs are mostly not conserved evolutionarily, they are heavily regulated suggesting their functional role. LncRNAs may function either as signal transducers, protein guides, or molecular scaffolds to regulate transcriptional and epigenetic events. Some LncRNAs perform these functions in cis- by modulating transcription of nearby genes (Dum) or act in trans- by modulating genes at multiple distant loci (MALAT1), while some can do both (HOTAIR).

Methods

ecCEBPα sequence

In the recent version of GENCODE, ecCEBPα is not annotated most likely due to an overlap with a protein-coding gene CEBPa on the same strand. We retrieved the complete ecCEBPα sequence from the human genome (hg19, chr19: 33298573-33303358) based on information reported in the work of Di Ruscio et al. ecCEBPα is approximately 4.8kb and it overlaps with the intronless CEBPa gene (~2.6kb) on the same strand. ecCEBPα does not share either the same transcription start site (TSS) or transcription end site (TES) with CEBPa, starting ~0.89kb upstream and ending ~1.46kb downstream of the CEBPa gene.
DNA methylation data processing
CpG methylation (Illumina 450K array) and CEBPα mutation data for 186 AML patients were retrieved from the Cancer Genome Atlas (TCGA: http://firebrowse.org). CpG methylation levels were measured in 307796 unique loci. We split all the AML patients into two groups based on CEBPα mutation status (13 patients with a CEBPα mutation and 173 patients without a mutation). We classified a CpG position as hypermethylated (HM) in patients with a CEBPα locus mutation if DNA methylation level was significantly increased in the case of a CEBPα mutation (t-test, FDR ≤0.05 and Δ-value ≥0.1) and all non-hypermethylated (NHM) CpG in the case of a CEBPα mutation were classified as non-hypermethylated CpGs (t-test, FDR >0.05 and |Δ-value| <0.1). As a result, we obtained 11955 HM and 261433 NHM CpGs.

Secondary structure and triplex prediction
As suggested in a previous study, unpaired RNA nucleotides are more likely to form triplexes with DNA. We predicted RNA secondary structure using RNAfold (V 2.4.14), from the Vienna suite using a cut-off for pairing probability (-c) of 0.1 [12,13]. To search for potential interactions between ecCEBPα and DNA target regions we used only unpaired nucleotides, while the nucleotides predicted to pair were replaced with ‘N’.

DNA target regions were defined as 100 nucleotides centered at each CpG. To predict ecCEBPα triplex formation with DNA target regions, we used Triplexator (V 1.3.2) [14], since it has higher accuracy of prediction [14, with the following optimization parameters suggested in 22: minimum length = 10 nucleotides, error rate = 20%, G-C content = 70%, and filter-repeats = off. Using these parameters out of 307796 unique CpG loci, we predicted 272131 loci with at least one triplex and 35715 without any. Among them, 10351 and 222105 potential triplex targets were predicted in the HM and NHM regions respectively.

To estimate the statistical significance of predicted triplexes we used Triplex domain finder (TDF v 0.12.3), which clusters RNA triplex-forming oligonucleotide (TFO) into DNA binding domains (DBD) [16]. Briefly, all 272131 CpG loci with at least one predicted triplex were taken as input target regions. By predicting triplexes in the background regions, TDF is capable of estimating the statistical significance of ecCEBPα binding between target regions and other non-target CpG regions. Since TDF allows only to mask regions in the genomic background rather than to select the background explicitly we had to prepare a special mask for the non-target regions. To do so we removed 35715 CpG loci with zero triplex predictions from the human genome using BEDTools subtract (BEDTools v2.29.2). TDF was implemented with a minimum triplex length (-l) of 10 nucleotides, an error rate (-e) of 20%, and (-f) to mask background loci in 100 random samplings (-n).

RNA:chromatin colocalization analysis
To validate the predicted interactions we used RNA:chromatin interactome obtained with iMARGI method capturing chromatin-associated RNA (caRNA) and their genomic interaction loci [17]. The data was downloaded from GEO (GSM3478205). The iMARGI dataset was mapped to the hg38 genome assembly. We used UCSC LiftOver to convert ecCEBPα sequence coordinates from hg19 to hg38 sequences [20]. We expanded the DNA coordinates of CpGs by 3.0kb nucleotides upstream and downstream. IntersectBed from BEDTools was used to check the co-location of predicted triplexes and experimentally validated interactions of ecCEBPα [21]. Fisher’s exact test was calculated for the number of confirmed ecCEBPα interactions between TDF and iMARGI data.

GO enrichment analysis
Finally, since Illumina 450K array probes are located close to genes, we performed functional enrichment using BiNGO (v 3.0.3) (binomial test) [20] to infer the biological significance of the genes potentially affected by ecCEBPα binding.

All statistical analyses were performed using R 4.0 or SciPy v1.5.1 library. Visualization was done in Cytoscape 3.2.0 [15] and Python 3.7. Code is available at https://zenodo.org/record/4385259.

Results
ecCEBPα forms triplexes with promoter regions and affects promoter methylation
The current study explores the potential of the lncRNA ecCEBPα in the modulation of global CpG methylation status in trans via direct interaction with DNA regions. ecCEBPα (extra coding CEBPα), reported in work by Di Russo et al. [22], is located on chromosome 19 and transcribed from the CEBPα locus (Figure 1a).

Mutations in CEBPα locus are a common feature of AML leading to whole genome hypermethylation (Figure 1b). Since TCGA is focused on protein-coding genes, all reported mutations are located within the CEBPα gene and could affect both CEBPα and ecCEBPα. To investigate if ecCEBPα could affect DNA methylation in trans, first we checked if it is capable of interaction via forming triple helices (triplexes) with its binding sites and if such interactions affect DNA methylation. We observed that regions capable of forming triplexes with ecCEBPα remain protected from global DNA hypermethylation observed in case of a mutation in a CEBPα locus (Figure 1c, Fisher’s exact test, p-value <0.001). Furthermore, the overall methylation profile of HL-60 cells with overexpressed ecCEBPα (Wang et al.) also have a protective effect on ecCEBPα’s binding sites in comparison to the rest of the genomic CpGs (Figure 1d, Fisher Test: p-val = 1.24E-09). This result suggests a negative relationship between DNMT access to promoter sites and ecCEBPα binding.

ecCEBPα binding is not affected by the mutation in the CEBPα locus
To investigate deeper the potential of ecCEBPα to form triplexes we used Triplex Domain Finder (TDF) - a triplex prediction tool that refines the resolution of predicted TFOs in RNA into DNA binding domains (DBD) and calculates the significance of the number of predicted triplexes for each DBD. Overall, 17 significant DBDs were identified within ecCEBPα, interspaced
between sequences 4086 and 4968 towards the 3’ end of the lncRNA (Figure 1e). These DBDs form triplexes with the majority of regions protected from hypermethylation in patients with a CEBPα mutation (Figure 1f, Extended data: Supplementary Table 1). The DBD region is located downstream from the CEBPα gene suggesting that ecCEBPα binding region is not affected by the mutation in the CEBPα locus. The predicted secondary structure of the ecCEBPα sequence showed that more than 95% of sequence positions from 4087-4987 (~0.5kb from CEBPα TES) (Figure 1e) were unpaired and potentially capable of forming triple helices with the target DNA region (Figure 2a).

Figure 1. (a) Schematic diagram for transcriptional products in the CEBPα locus; CEBPα is represented by an orange arrow and ecCEBPα is represented by the blue line. (b) Global change in DNA methylation. (c) Difference in DNA methylation levels between patients with and without CEBPα mutation in the regions of ecCEBPα predicted binding (non parametric t-test, p-value >1E-10). (d) Number of DNA Triplex Target Site (y-axis) and a location of the corresponding TFO on the ecCEBPα (TFO: RNA; x-axis). (e) Number of Triplex Target sites per TFO predicted for NHM and HM CpG regions.

eCebpa interacts with predicted binding sites
Since we use relatively relaxed thresholds for triplex prediction, we decided to validate the predicted RNA:DNA triplexes using experimental data obtained with the iMARGI method, allowing detection of RNA-chromatin interactions. We identified 157 ecCEBPα contacts within the iMARGI dataset and 29 of them contained predicted triplexes. Altogether, these 29 iMARGI interactions were made up of 182 predicted TTS (Fisher’s exact test, p-value < 2.2E-16) located in cis and in trans on 14 chromosomes (Figure 2b). Chromosomes 19 (the native chromosome for ecCEBPα) are accounted for by all predictions. Since these ecCEBPα contacts have been
experimentally validated, we suggested that they are the most reliable regions where ecCEBPA might have a protective effect from DNA methylation. Overall, a mean methylation delta score of experimentally validated binding sites is significantly lower than that of the non-binding sites (p-value < 1E-09) (Figure 2c).

Genes protected from methylation by ecCEBPa are involved in transcription factor activities

We performed gene ontology analysis on the genes located nearby 182 ecCEBPa triplexes supported by iMARGI. Key gene ontology categories such as nucleic acid binding and transcription factor activities were enriched among putative ecCEBPa targets (Figure 2d). Representative genes among transcription factors include MLF2, SUV39H2, RBM5, UBTF, and among sequence-specific DNA binding proteins include POU2F2, MED12L, and DNASE1L1. The enrichment in transcription factors (TF) suggests that triplex formation may represent a possible mechanism employed by ecCEBPa to regulate TF methylation and as a consequence, their expression. Furthermore, out of the 33 genes we identified to be targets of ecCEBPA, 16 genes (ICAM1, PDXK, etc.) are clearly related to hematopoiesis and various leukemias. A summary of the genes and their function is provided in Table 1.

Discussion

Unlike other forms of cancers, AML progression is often mutation-independent but may be explained by altered epigenetic regulation, DNA methylation specifically\(^1\). In this study, we elucidate a putative mechanism for the regulation of DNA methylation by ecCEBPa. In a previous study, ecCEBPa, which accompanies the transcription of CEBPa on the same locus, was shown to protect the promoter of CEBPa from DNA methylation leading to active expression\(^2\). We speculated that ecCEBPa might perform a similar function in trans. We demonstrated that ecCEBPa-DNA triplex formation might provide the molecular basis of this interaction. ecCEBPa binding presumably protects the region from genome-wide hypermethylation induced by CEBPa mutation in AML patients. Diruscio et al. confirmed that the most mutations in CEBPa do not influence the expression of ecCEBPa but rather, its ability of the RNA to fold properly. We suggest that the inability to fold
## Table 1. AML specific function of nearby genes that are protected by ecCEBPA.

| Gene     | Role                                                                 | Reference                                                                                   |
|----------|----------------------------------------------------------------------|---------------------------------------------------------------------------------------------|
| FLVCR1   | Alternative splicing disrupts of FLVCR1 disrupts erythropoiesis in Diamond-Blackfan anemia. (Rey et al. 2008) | https://haematologica.org/article/view/5065                                                 |
| PSMC6    | Identified as part of a novel cluster for classifying AML risk and predicting outcomes. (Wilson et al. 2006) | https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1895492/                                       |
| CMIP     | Identified as one of the top genes expressed after rhIGFBP7 treatment (to suppress clonal growth) of AML cases. (Verhagen et al. 2018) | https://www.cell.com/cell-reports/pdfExtended/S2211-1247(18)31838-2                          |
| GNA13    | A crucial signalling component of the GPR84/Beta-Catenin Signaling Axis in AML Stem Cells. (Dietrich et al. 2014) | https://ashpublications.org/blood/article/124/21/3577/97803/GNA13-a-Novel-Component-of-the-GPR84-Beta-Catenin |
| ICAM1    | Responsible for migration and adhesion of myeloid cells in hyperleukocytic AML. (Zhang et al. 2006) | https://onlinelibrary.wiley.com/doi/full/10.1111/j.1365-2257.2006.00784.x                    |
| PDE4C    | Its mutation/absence leads to suppression of apoptotic response in myelogenous symptoms. (Lerner and Epstein 2006) | https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1383661/                                       |
| ANKRD27  | Identifed as part of a novel cluster of competing endogenous RNA that predict AML survival and prognosis. (Wang et al. 2020) | https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7432272/                                       |
| PDXK     | Responsible for Vitamin B6 addiction of AML cells. (Chen et al. 2020) | https://pubmed.ncbi.nlm.nih.gov/31935373/                                                  |
| NLRP3    | The NLRP3 inflammasome is upregulated as part of the stress response in AML cells. (Jia et al. 2017) | https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5754918/                                       |
| PNPLA3   | Responsible for elevated transaminases in lymphoblastic anaemia. (Bruschi et al. 2017) | https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5683790/                                       |
| CELSR1   | Potentially contains mutations that contribute to patient specific mutations that are responsible for the origination of AML. (Skoczen et al. 2019) | https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3407563/                                       |
| PFKFB4   | Responsible for survival of acute monocytic myeloid leukemia by suppression of apoptosis. (Wang et al. 2020) | https://pubmed.ncbi.nlm.nih.gov/32299611/                                                  |
| COL7A1   | Contains somatic variations that are described to be potentially pathogenic in ALL patients. (Skoczen et al. 2019) | https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6978700/                                       |
| MED12L   | Regulates hematopoietic stem cell (HSC) specific enhancers. It loss leads to loss of HSC stemness. (Aranda-Orgilés et al. 2016) | https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5268820/                                       |
| GALNTL5  | Contains deleterious deletions that lead to copy number alteration in core binding factors of both adult and pediatric AMLs. (Kühn et al. 2012) | https://ashpublications.org/blood/article/119/10/e67/29576/High-resolution-genomic-profiling-of-adult-and   |
| TAZ      | Controls AML stemness and differentiation by modulating toll-like receptor (TLR) signaling. (Seneviratne et al. 2019) | https://pubmed.ncbi.nlm.nih.gov/30930145/                                                  |
properly even in overexpressed cases may affect its structural ability to bind its targets. ecCEBPα contains a TFO/DBD-rich region at its 3’ end, with low pairing probability, suggesting that it is capable of triplex formation. Several of the predicted ecCEBPα binding sites, which include transcription factors such as MLF2, SUV39H2, RBM5, UBTF, and sequence-specific DNA binding proteins include POU2F2, MED12L, and DNASE1L1, were supported by experimental iMARGI RNA:chromatin interactions data.

Currently, the understanding of lncRNA function and mechanisms of action is limited to a few dozen well-annotated lncRNA transcripts. A few functional characterization attempts are based on the ‘guilt by association’ hypothesis, which may not resonate well with the ability of lncRNA to interact in trans. As thoroughly reviewed previously, lncRNAs such as ecCEBPα, Dum, Dalí, Dacor1, and LINC RNA-P21 interact with DNA in trans to regulate DNA methylation. The results presented herein further demonstrate that triplex formation between ecCEBPα and CpG containing DNA regions could indeed be regulatory and protect CpG sites from DNMT activity.

Unfortunately, RNA:chromatin interaction protocols are relatively new and the data is available only for a few cell types. Since RNA:chromatin interactions are highly cell-type specific and lowly expressed, it is not surprising that we could validate only a few of the predicted interactions. Nevertheless, based on our results, we suggest a model of potential ecCEBPα chromatin interaction in trans (Figure 2e). In this model, ecCEBPα uses its unpaired regions to directly bind to specific DNA sequences by forming triplexes and in this way prevents DNA methylation in the region of binding. ecCEBPα binding to distant regions could be mediated either by 3-dimensional chromatin organization or which brings them close to ecCEBPα.

Recent studies have observed that promoter or transcription start sites (TSS) regions, which tend to be rich in CpG dinucleotides, are TTS-rich and potential triplex-forming hotspots. Through functional enrichment analysis, we observed that transcription factors might be preferential targets of ecCEBPα. Interestingly, previous studies have shown that the suppression of a myeloid leukemia factor (MLF2), an oncogene in breast cancer and myeloid leukemia, as well as UBTF which controls rDNA expression, significantly contributes to cancers upon promoter hypermethylation. The suppressor of variegation 3-9 homolog 2 (SUV39H2), a histone-lysine-N-methyltransferase which regulates the hypermethylation H3K9 has also been reported to indirectly influence over 450 promoters in AML. Having in mind that ecCEBPα is transcribed from CEBPα locus - a key transcription factor of hematopoiesis - this lncRNA could participate in the formation of a hub in the hematopoiesis regulatory network.

**Conclusion**

In conclusion, we have shown that ecCEBPα could serve as a trans-acting regulatory agent protecting its binding sites from genome-wide CpG methylation, and its dysregulation could contribute to aberrant methylation profile in AML patients. These results suggest a novel regulatory mechanism for ecCEBPα as a modulator of DNA methylation through triplex formation providing a foundation for sequence-specific engineering of RNA for regulating methylation of specific genes.

**Data availability**

**Underlying data**

**Extended data**

**Data availability**

This project contains the following underlying data:

- DNA_BINDING_DOMAINS_ID.tsv
- predicted_secondary_structure_of_ecCEBPA.fa
- probes.csv (main data)

Code used for analysis available from: https://github.com/josoga2/eccebp-alpha-project/tree/f1000

Archived code as at time of publication: https://doi.org/10.5281/zenodo.4385259

This project contains the following extended data:

- Supplementary Table 1: Summary table of DNA binding domains (DBD), the counts of target regions within the genome and statistical analysis.
- ecCEBPα secondary structure prediction with RNAplfold

Data and code are available under the terms of the Creative Commons Attribution 4.0 International license (CC-BY 4.0).
Open Peer Review

Current Peer Review Status: 

Version 2

Reviewer Report 15 September 2021

https://doi.org/10.5256/f1000research.65190.r93095

© 2021 Singh A. This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Amrita Singh

CSIR Institute of Genomics and Integrative Biology (CSIR-IGIB), New Delhi, Delhi, India

All the major comments raised by the reviewer have been answered thoroughly by the authors. The additional methylation analysis of HL-60 datasets in both normal and overexpressed ecCEPBα is highly appreciated. Moreover, re-labelling of the plots as highlighted previously has made it easier to comprehend them. However, there are few minor technical remarks, both figures 1 and 2 do not have the legend for the updated figure 1(d) and 2(c), and accordingly, the subsequent legends labelling needs to be modified.

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Structure function relationship of noncoding RNA, with major focus on LncRNA function via triple helical structures.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Version 1

Reviewer Report 10 June 2021

https://doi.org/10.5256/f1000research.31133.r81320

© 2021 Demidov O. This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Oleg N. Demidov
INSERM U1231, Faculty of Medicine and Pharmacy, University of Bourgogne-Franche Comté, Dijon, France

The manuscript by Ogunleye AF and co-authors describes a novel phenomenon of IncRNA ecCEBP α-dependent regulation of methylation in several genes. The authors analyzed the open datasets and showed that mutations in CEBPA gene (coding both, CEBPA transcriptional factor and IncRNA ecCEBPα) changed the global CpG methylation status in AML cells. The manuscript is nicely written, conclusions are logically justified. This work will be interesting to the readers in the epigenetic regulation and cancer research field.

Still, I have several concerns about the analysis. There are a lot of predicted triplex target regions for ecCEBPα genome-wide. Some of them could be false-positive predictions. Is there any way to confirm the observed effect only of experimentally validated contacts? Or in regions where predictions are most reliable?

Also, it is unclear other genes, not only transcription factors, predicted to be regulated by ecCEBPα are related to hematopoiesis. It would be helpful to have at least some hypotheses in the discussion. If there is some link found between ecCEBPα targets and regulation of hematopoiesis, it would really strengthen the conclusions of the work. Probably, KEGG pathway enrichment analysis may help with that.

**Minor comments:**

It is unclear on the figure with a secondary structure of ecCEBPα where the triplex-forming region is located.

**Is the work clearly and accurately presented and does it cite the current literature?**
Yes

**Is the study design appropriate and is the work technically sound?**
Yes

**Are sufficient details of methods and analysis provided to allow replication by others?**
Yes

**If applicable, is the statistical analysis and its interpretation appropriate?**
Yes

**Are all the source data underlying the results available to ensure full reproducibility?**
Yes

**Are the conclusions drawn adequately supported by the results?**
Yes

**Competing Interests:** No competing interests were disclosed.

**Reviewer Expertise:** Models of human diseases, cancer research, fibrosis, paging
I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Author Response 05 Aug 2021

ADEWALE OGUNLEYE, Moscow Institute of Physics and Technology, Moscow, Russian Federation

Response to Reviewer Two

Comment one: There are a lot of predicted triplex target regions for ecCEBPA genome-wide. Some of them could be false-positive predictions. Is there any way to confirm the observed effect only of experimentally validated contacts? Or in regions where predictions are most reliable?

Response one: We are grateful to the reviewer for this suggestion. We performed a validation for the most reliable predicted triplexes using only experimentally validated ecCEBPA:DNA contacts. Overall, a mean methylation delta score of only 0.01 (p-value < 1E-09) was observed which further emphasizes the protective effect of ecCEBPA. The following text and images were added to the main body of the paper:

“Since these ecCEBPA contacts have been experimentally validated, we suggested that they are the most reliable regions where ecCEBPA might have a protective effect from DNA methylation. Overall, a mean methylation delta score of experimentally validated binding sites is significantly lower than that of the non-binding sites (p-value < 1E-09) (Figure 2C).”

New figure panels (1&2) here: https://drive.google.com/file/d/1iynm-t0DfGkxLF5vsumG4iCcfo2u2xCJ/view?usp=sharing

Comment two: Also, it is unclear if other genes, not only transcription factors, predicted to be regulated by ecCEBPA are related to hematopoiesis. It would be helpful to have at least some hypotheses in the discussion. If there is some link found between ecCEBPA targets and regulation of hematopoiesis, it would really strengthen the conclusions of the work. Probably, KEGG pathway enrichment analysis may help with that.

Response two: We thank the reviewer for this important comment. Indeed, we discovered that other non-transcription factor genes have links to acute myeloid leukemia via literature mining. Unfortunately, this could not be enriched through KEGG. We suggest including a supplementary table that details the different roles of the identified genes.

In the main text, we added:

“Out of the 33 genes we identified to be targets of ecCEBPA, 16 genes (ICAM1, PDXK, etc.) are clearly related to hematopoiesis and various leukemias. A summary of the genes and their function is provided in Table 1.”

Table 1: AML specific function of nearby genes that are protected by ecCEBPA (added to the paper).

Competing Interests: No competing interests were disclosed.
The article has addressed an important question of the ability of a lncRNA to act in trans and regulate a multitude of genes, which might be affecting the progression of acute myeloid leukemia. The authors have also highlighted the possibility of a triplex structure for defining the binding of lncRNA at its target gene CpG region. However, there are certain parts wherein the authors haven’t been able to convey their points properly.

1. In Figure 1(b), a global DNA hypermethylation has been represented in AML patients with CEBPα mutation, but the numbers of the HM and NHM sites in the methods section do not reflect the same. It shows higher sites for NHM, than the HM sites, kindly check if some mislabeling is there at the authors part.

2. Also, the reviewer has failed to understand Figure 1(c). In both the x-axis and the side legend, binding and non-binding have been depicted, but which one is the AML patients with and without mutation isn’t clear from the figure.

3. There are few other key points that were not clear;

(a) what is the level of ecCEBPα expression in AML patients with and without mutation.

(b) Additionally, if the expression of the lncRNA remains the same in both cases, how does one explain the increase in binding of the lncRNA and subsequent higher NHM sites, in the case of AML patients with CEBPα mutation?

(c) Moreover, a previous report (Di Ruscio et.al/) on ecCEBPα had also analyzed genome-wide methylation, in which they report that methylation levels remain unchanged even when ecCEBPα was overexpressed. This is in contrast with the major theme of the paper i.e., methylation of genes in trans is affected by the ecCEBPα. The authors should comment on this in the discussion part.

References
1. Di Ruscio A, Ebralidze AK, Benoukraf T, Amabile G, et al.: DNMT1-interacting RNAs block gene-specific DNA methylation. Nature. 2013; 503 (7476): 371-6 PubMed Abstract | Publisher Full Text

Is the work clearly and accurately presented and does it cite the current literature?
Partly
Is the study design appropriate and is the work technically sound?
Yes

Are sufficient details of methods and analysis provided to allow replication by others?
Partly

If applicable, is the statistical analysis and its interpretation appropriate?
I cannot comment. A qualified statistician is required.

Are all the source data underlying the results available to ensure full reproducibility?
Yes

Are the conclusions drawn adequately supported by the results?
Partly

**Competing Interests:** No competing interests were disclosed.

**Reviewer Expertise:** Structure function relationship of noncoding RNA, with major focus on LncRNA function via triple helical structures.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

---

**Author Response 05 Aug 2021**

**ADEWALE OGUNLEYE**, Moscow Institute of Physics and Technology, Moscow, Russian Federation

**Response to Reviewer One**

**Comment One:** In Figure 1(b), a global DNA hypermethylation has been represented in AML patients with CEBPa mutation, but the numbers of the HM and NHM sites in the methods section do not reflect the same. It shows higher sites for NHM, than the HM sites, kindly check if some mislabeling is there at the authors part.

**Response Two:** We believe that there is a confusion here. The plot was correct and reflected the change in DNA methylation in all CpG positions including those that were not statistically differentially methylated. We counted hypermethylated CpG (t-test, FDR ≤0.05 and Δ-value ≥0.1) and those that were not hypermethylated (t-test, FDR >0.05 and |Δ-value| <0.1). In this way, we refer to non-hypermethylated (NHM) to any CpG with DNA methylation change less than 0.1, which includes either not changing or hypomethylated and not only hypomethylated. To avoid this confusion we suggest replacing the plot with the one below. We specified a 0.1 mid-point to aid visualization. (Figure 1b).

**Comment Two:** Also, the reviewer has failed to understand Figure 1(c). In both the x-axis and the side legend, binding and non-binding have been depicted, but which one is the AML patients with and without mutation isn't clear from the figure.
Response Two: Each violin in Figure 1(c) is a representation of the differential methylation score between CpGs of AML patients with and without CEBPa mutation. The blue violin represents all CpGs that are located in ecCEBPA binding sites, while the orange violin represents CpGs that do not bind with ecCEBPA. To provide clarity, we corrected the labels on the X and Y axes and added explicit explanation in the legend of the figure (Figure 1c).

Difference in methylation = Mean methylation of CpGs in AML patients with CEBPA mutation - Mean methylation of CpGs in AML patients without CEBPA mutation.

Comment 3: There are few other key points that were not clear;
(a) what is the level of ecCEBPα expression in AML patients with and without mutation.
(b) Additionally, if the expression of the IncRNA remains the same in both cases, how does one explain the increase in binding of the IncRNA and subsequent higher NHM sites, in the case of AML patients with CEBPα mutation?
(c) Moreover, a previous report (Di Ruscio et.al1) on ecCEBPα had also analyzed genome-wide methylation, in which they report that methylation levels remain unchanged even when ecCEBPα was overexpressed. This is in contrast with the major theme of the paper i.e., methylation of genes in trans is affected by the ecCEBPα. The authors should comment on this in the discussion part.

Response:
(a&b): It is relatively difficult to estimate the expression level of ecCEBPA since this gene is not in the annotation used by TCGA. The raw data is not freely available in TCGA. ecCEBPA gene also overlaps with the CEBPA gene making the estimation of expression of each of the genes even more complicated. Since the mutation in AML patients happens in the body of CEBPA gene, we do not expect a significant change in the expression of either CEBPA or ecCEBPA.

We explained the relationship between ecCEBPA binding and methylation in the first result section (Figures 1c,d&e). We suggested that the mutation(s) in CEBPA does not affect the expression of the IncRNA, but rather the ability to bind its targets.

We further emphasize this point in the paper with this line (discussion): “Di Ruscio et al. confirmed that most mutations in CEBPα do not influence the expression of ecCEBPα but rather, its ability of the RNA to fold properly. We suggest that the inability to fold properly even in overexpressed cases may affect its structural ability to bind its targets.”

(c): We are very grateful to the reviewer for this valuable comment. In response to this, we retrieved the DNA methylation data (RRBS) for wild-type HL-60 cells (ENCODer Dataset) and overexpressed HL-60 cells (Di Ruscio et al). We then calculated the methylation difference between the two cell states and compared it between the ecCEBPA binding sites versus the non-binding sites. Non ecCEBPA binding sites were more methylated in comparison to ecCEBPA binding sites (Fisher Test: p-val = 1.24E-09). Bearing in mind that ecCEBPA targets are located genome-wide, it is tempting to suggest that “enforced overexpression” (which was achieved with the R1 variant of ecCEBPA; comprise of downstream ecCEBPA sequences) of ecCEBPA strongly protects local CpGs from methylation while the rest of the genome gain some methylation.
We added the following statement to the results in the main text:
"Furthermore, the overexpression of ecCEBPA in HL-60 cells (Wang et al) lead to ecCEBPA binding sites stay unmethylated while non ecCEBPA binding sites gain methylation (Fisher Test: p-val = 1.24E-09), suggesting a protective effect on ecCEBPA's binding to its target locations (Fig 1d)."

**Competing Interests:** No competing interests were disclosed.

---

## Comments on this article

**Version 1**

Reader Comment 14 Apr 2021

Olumide Inyang, moscow institute of physics and technology (MIPT), Russian Federation

This report gave a clear understanding of ecCEBPα as a genome-wide epigenetic modulator via triple-helix stacks for RNA(engineered) in regulating methylation of specific genes...great work from the authors.

**Competing Interests:** No competing interests were disclosed.

---

The benefits of publishing with F1000Research:

- Your article is published within days, with no editorial bias
- You can publish traditional articles, null/negative results, case reports, data notes and more
- The peer review process is transparent and collaborative
- Your article is indexed in PubMed after passing peer review
- Dedicated customer support at every stage

For pre-submission enquiries, contact research@f1000.com