PHF6 interacts with the LMO2 complex in T-cell acute lymphoblastic leukemia

Vesna S. Stanulović,[1,2*] Sarah Binhassan,[1,2*] Budoor A. Jaber,[1] Shimaa Alazmi,[1,2] Fatma M.B. Saleman,[1,2] Sandeep Potluri,[1,2] Guy Pratt,[1,3] Christian Ludwig,[4] Douglas G. Ward[1] and Maarten Hoogenkamp[1,2]

[1]Department of Cancer and Genomics Sciences, School of Medical Sciences, College of Medicine and Health, University of Birmingham; [2]Birmingham Center for Genome Biology, University of Birmingham; [3]Center for Clinical Hematology, Queen Elizabeth Hospital Birmingham and [4]Department of Metabolism and Systems Sciences, School of Medical Sciences, College of Medicine and Health, University of Birmingham, Birmingham, UK

**VSS and SB contributed equally as first authors.*

Correspondence: M. Hoogenkamp
m.hoogenkamp@bham.ac.uk

Received: December 5, 2024.
Accepted: June 24, 2025.
Early view: July 10, 2025.

https://doi.org/10.3324/haematol.2024.287124

Bioinformatical analysis

NGS sequencing analysis was performed on usegalaxy.org, usegalaxy.eu (1).

RNAseq-reads were mapped to the mouse genome (GRCm38/mm10) using HISAT2-FeatureCounts-Deseq2 workflow and annotation genecode v21h (2). Adjusted p-value <0.05 was used to identify significantly differentially expressed genes. Heat maps and the hierarchical clustering, based on Pearson correlation with average linkage clustering, were computed by MultiExperiment Viewer v4.9.0. Gene ontology (GO) enrichment was performed using DAVID 6.8 and GREAT (3, 4). GO terms with p<0.05 were considered significant; redundant terms were removed.

ChIPseq reads were mapped to hg39 using HISAT2. Peaks were called using MACS2 (5-7). Peaks were considered overlapping if there was any overlap of the peak coordinates. For ChIP-RNAseq integration we used gene expression files that contained all genes with gene count >1 in at least one sample and the nearest 5'- and 3'-genes were selected based on the genomic coordinates of the ChIPseq peaks. Overlays were generated using EaSeq (8). Nearest 5' and 3' genes to PHF6 peaks were identified using genecode v21h.

Transcriptional regulatory architecture analysis was conducted using Cytoscape (version 3.10.2)(9) with the stringApp (version 2.1.1)(10), and the DoRothEA human regulon database (version 1.17.0)(11). The Prefuse Force Directed layout (12) was utilized to visualize the network.

References

1. Afgan E, Baker D, Batut B, van den Beek M, Bouvier D, Cech M, Chilton J, Clements D, Coraor N, Gruning BA, Guerler A, Hillman-Jackson J, Hiltemann S, Jalili V, Rasche H, Soranzo N, Goecks J, Taylor J, Nekrutenko A, Blankenberg D. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. Nucleic Acids Res. 2018 Jul 2;46(W1):W537-W544. doi:10.1093/nar/gky379. Cited in: Pubmed; PMID 29790989.

2. Kim D, Langmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory requirements. Nat Methods. 2015 Apr;12(4):357-60. doi:10.1038/nmeth.3317. Cited in: Pubmed; PMID 25751142.

3. Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc. 2009;4(1):44-57. doi:10.1038/nprot.2008.211. Cited in: Pubmed; PMID 19131956.

4. McLean CY, Bristor D, Hiller M, Clarke SL, Schaar BT, Lowe CB, Wenger AM, Bejerano G. GREAT improves functional interpretation of cis-regulatory regions. Nat Biotechnol. 2010 May;28(5):495-501. eng. Epub 2010/05/04. doi:10.1038/nbt.1630

nbt.1630 [pii]. Cited in: Pubmed; PMID 20436461.

5. Feng J, Liu T, Qin B, Zhang Y, Liu XS. Identifying ChIP-seq enrichment using MACS. Nat Protoc. 2012 Sep;7(9):1728-40. doi:10.1038/nprot.2012.101. Cited in: Pubmed; PMID 22936215.

6. Khan A, Mathelier A. Intervene: a tool for intersection and visualization of multiple gene or genomic region sets. BMC Bioinformatics. 2017 May 31;18(1):287. doi:10.1186/s12859-017-1708-7. Cited in: Pubmed; PMID 28569135.

7. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 2010 Mar 15;26(6):841-2. doi:10.1093/bioinformatics/btq033. Cited in: Pubmed; PMID 20110278.

8. Lerdrup M, Johansen JV, Agrawal-Singh S, Hansen K. An interactive environment for agile analysis and visualization of ChIP-sequencing data. Nat Struct Mol Biol. 2016 Apr;23(4):349-57. doi:10.1038/nsmb.3180. Cited in: Pubmed; PMID 26926434.

9. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 2003 Nov;13(11):2498-504. doi:10.1101/gr.1239303. Cited in: Pubmed; PMID 14597658.

10. Doncheva NT, Morris JH, Gorodkin J, Jensen LJ. Cytoscape StringApp: Network Analysis and Visualization of Proteomics Data. J Proteome Res. 2019 Feb 1;18(2):623-632. The authors declare no competing financial interest. Epub 20181205. doi:10.1021/acs.jproteome.8b00702. Cited in: Pubmed; PMID 30450911.

11. Garcia-Alonso L, Holland CH, Ibrahim MM, Turei D, Saez-Rodriguez J. Benchmark and integration of resources for the estimation of human transcription factor activities. Genome Res. 2019 Aug;29(8):1363-1375. Epub 20190724. doi:10.1101/gr.240663.118. Cited in: Pubmed; PMID 31340985.

12. Heer J, Card SK, Landay JA. prefuse: a toolkit for interactive information visualization. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems; 2005 Portland, Oregon, USA. Association for Computing Machinery; 2005 p. 421–430. Available from: https://doi.org/10.1145/1054972.1055031.

Supplemental Figure Legends

Figure S1. Immunophenotypes of the T-ALL cell lines, with their EGIL classification, used within this study, as described by Sandberg *et al.* (2007).

Figure S2. Frequency of occurrence of PHF6 mutations in different oncogenic subtypes of T-ALL. Data reported for 1309 T-ALL cases by Pölönen *et al.* (2024) was analysed. Chi square tests were performed, which showed that PHF6 mutations were statistically significantly enriched ($p<0.05$) in their defined HoxA9 TCR, TLX1, and TLX3 subtypes and underrepresented in the KMT2A, MLLT10, TAL1 αβ-like, and TAL1 DP-like subtypes, compared to All. ns = not significant.

Figure S3. (A) Table showing the number of PHF6, TAL1, LMO2, LDB1 and GATA2 peaks found in ARR, DU.528, HSB2 and CCRF-CEM cells and the number of their overlaps. Overlap between at least two of the ChIPseq peaks are coloured according to the number with low numbers in red and highest in green. **(B)** Table showing number and percent of overlapping peaks between PHF6 and any other of the investigated proteins for each T-ALL cell line.

Figure S4. Enrichment for the TAL1, LMO2, LDB1, GATA2 and PHF6 ChIP at the Runx1 intronic enhancer and the negative control region Chr.18, relative to the input control. Data points are the mean of at least three independent samples measured in duplicate ± StDev.

Figure S5. Heat maps showing overlay of the PHF6, LMO2, TAL1, LDB1, GATA2 ChIPseq results. Matrix was derived from the coordinates of the LMO2 peaks ±2 kb around the centre of the peaks, ranked according to the score of the intensity. Aligned reads from PHF6, LMO2, TAL1, LDB1, GATA2 ChIPseq were overlayed and the intensity of pixels colour corelates with the number of reads found at each genomic position. The colour bars positioned below the heat maps show the relation between the binding intensity and the colour.

Figure S6. Clustered heatmap of the Pearson Correlation analyses for the PHF6, TAL1, LMO2, GATA2 and LDB1 ChIPseq bound DNA elements in ARR, DU.528, HSB2 and CCRF-CEM cells. Dendrogram on the left illustrates the hierarchical relationships between the samples. Colour bar shows the relation between the colour and the correlation value.

Supplementary tables are available in separate Excel Files

Table S1. List of primer sequences and antibodies used.

Table S2. Mass spectrometry results. “T-ALL_LMO2_IP specific” lists the proteins that were identified for each cell line, “Combined IgG” lists all proteins that were identified in the IgG control samples, and “Common in 3 T-ALL” shows proteins that were identified in three out of four samples after subtraction of the proteins from the IgG controls.

Table S3. Analysis of ADHC PHF6 peaks. “ADHC_PHF6_peaks” shows the genomic location of the peaks, “Associated_genes” shows the genes that associate with the peaks according, and “David_GeneOntology” show the gene ontology results.

Table S4. Integration of ADHC-TLGLP ChIPseq peaks with RNAseq. “ADHC_TLGLPpeaks_FPKM” lists genomic coordinates of the ADHC-TLGLP peaks and the nearest expressed genes in 5’ and 3’ direction with their FPKM values, “Associated_genes” lists all the identified genes that were used for gene ontology analysis, and “David_GeneOntology” shows the gene ontology results.

Table S5. Integration of ADHC-TLGL ChIPseq peaks with RNAseq. “ADHC_TLGL_peaks_FPKM” lists genomic coordinates of the ADHC-TLGL peaks and the nearest expressed genes in 5’ and 3’ direction with their FPKM values, “Associated_genes” lists all the identified genes that were used for gene ontology analysis, and “David_GeneOntology” shows the gene ontology results.

Table S6. Analysis of MJT1T2 PHF6 peaks. “MJT1T2_P6_peaks” shows the genomic location of the peaks, “Associated_genes” shows the genes that associate with the peaks according, and “David_GeneOntology” show the gene ontology results.

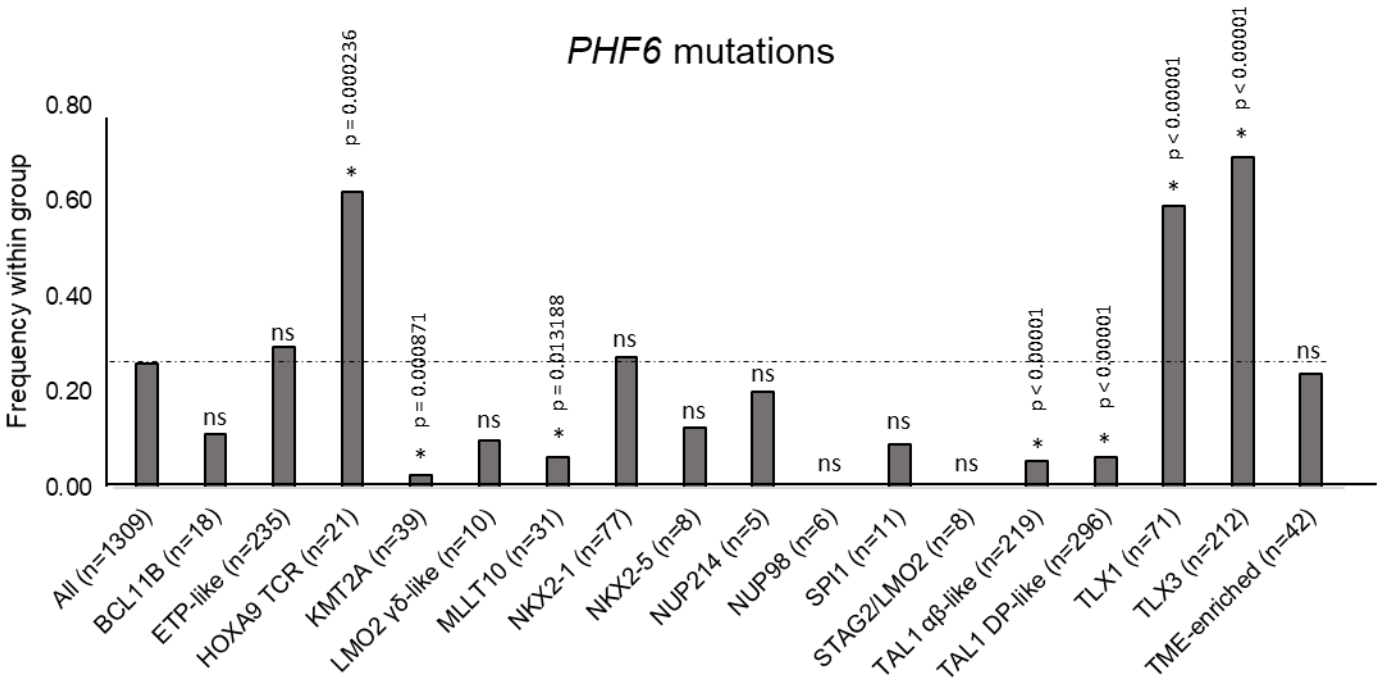
Table S7. RNAseq data for PHF6$^{-/-}$ ARR and Cas9 control ARR cells. “DEG” lists the statistically significantly differentially expressed genes, “GO_all” shows the gene ontology results for these genes, “higher_ARR” and “higher_PHF6KO” list the genes that are expressed at a higher level in the indicated cells. “GO_higher_ARR cells” and “GO_higher_PHF6KO” list the respective gene ontology identified by DAVID.

Table S8. Integration of ARR-PHF6 peaks which overlap with at least one of the other LMO2-associated factors and the nearest 5‘ and 3’ genes. “PHF6anyTLLG_RNAseq” lists genomic coordinates of the peaks and the nearest expressed genes in 5’ and 3’ direction with their FPKM values, “P6anyTLLG_padj005ARR” lists the identified genes that were significantly differentially expressed, “P6anyTLLG_genesUPinP6KO” shows the genes that were upregulated in PHF6$^{-/-}$ ARR cells, and “P6anyTLLG_genesDOWNinP6KO” lists the genes that were downregulated in PHF6$^{-/-}$ ARR cells

Figure S1

	ARR (T I/ETP)	DU.528 (T II)	HSB2 (T III)	CCRF-CEM (T IV)	Molt-4 (T III)	Jurkat (T IV)
CD34	+	-	-	-	+	-
CD7	+	+	+	+	+	+
CD5	-	+	+	+	+	+
CD1a	-	-	+	+	+	+
CD3	-	-	-	+	-	+
CD4	-	-	-	+	+	+
CD8	-	-	-	-	+	-

Figure S2



Pölönen et al. (2024) data set of 1309 cases

Figure S3

A

PHF6	x					x	x	x	x	x	x	x	x	x	x	x	x	x		x
LMO2		x				x				x	x	x			x	x	x		x	x
TAL1			x				x			x			x	x	x	x		x	x	x
LDB1				x				x			x		x		x	x		x	x	x	x
GATA2					x				x			x		x	x		x	x	x	x	x

ARR	846	10088	11326	6727	36864	39	18	322	218	32	11	81	0	12	1447	12	421	66	8	2338	1775
DU.528	1474	8592	7350	5800	2310	11	20	43	19	35	3	6	0	8	154	40	132	7	4	3627	986
HSB2	7644	13815	5087	3752	9258	53	27	79	120	55	5	30	2	15	279	15	219	18	0	4405	994
CCRF-CEM	3835	26045	3079	11397	8403	198	20	199	194	127	13	256	0	26	503	34	819	141	9	5889	4292

B

	PHF6 Peaks	Shared PHF6 Peaks	% of shared PHF6 peaks
ARR	6178	5308	69.42
DU.528	4984	2942	44.79
HSB2	6004	9555	68.45
CCRF-CEM	11989	10666	64.43

Figure S4

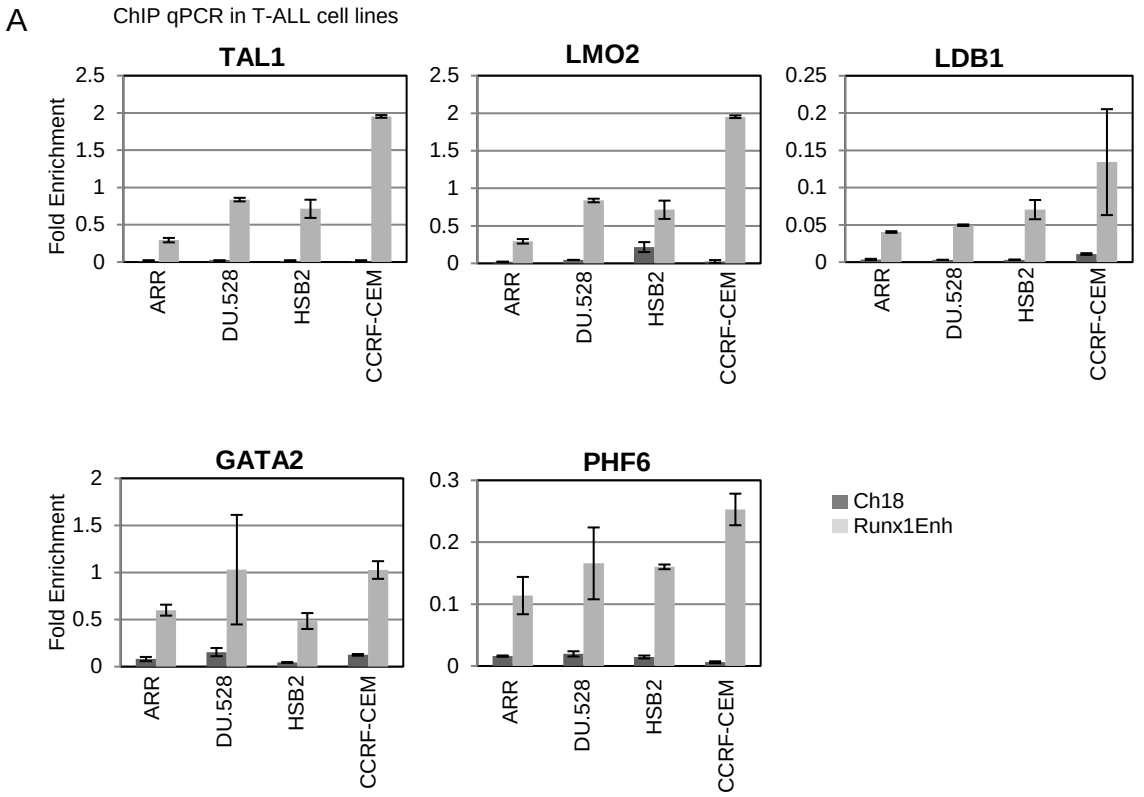

A
ChIP qPCR in T-ALL cell lines
TAL1
LMO2
LDB1
GATA2
PHF6
Fold Enrichment
ARR
DU.528
HSB2
CCRF-CEM
Ch18
Runx1Enh

Figure S5

A

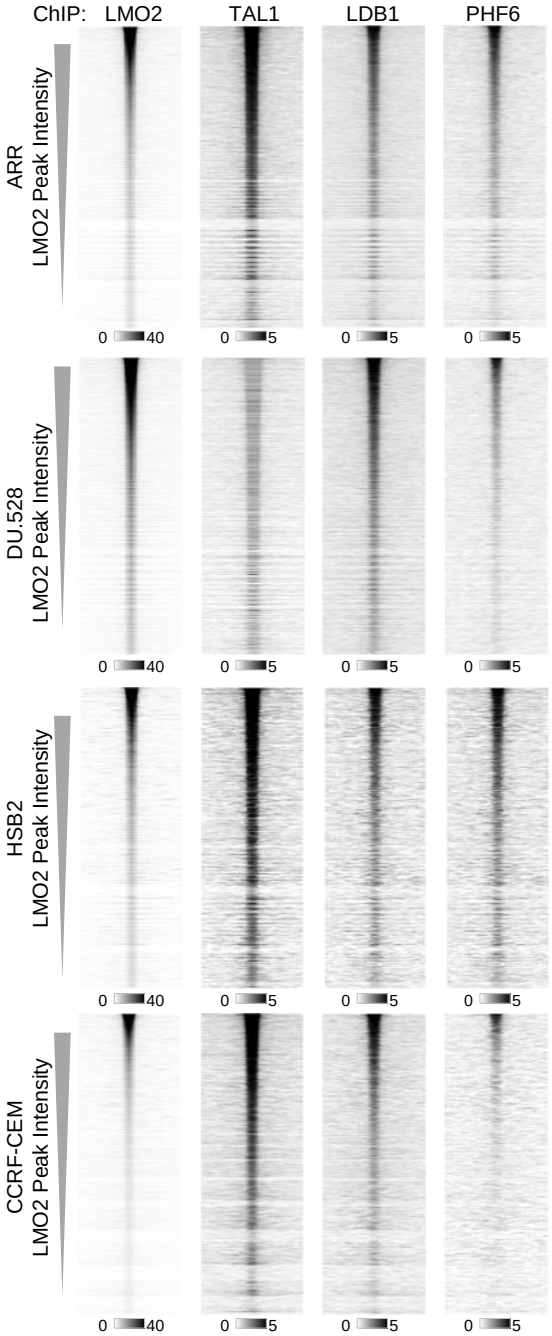
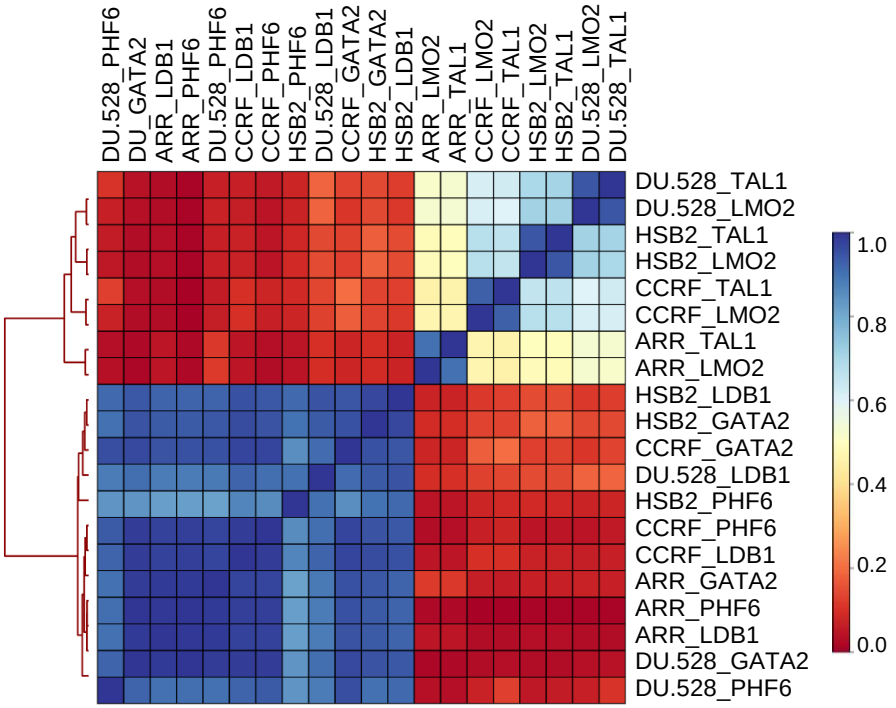

ChIP: LMO2 TAL1 LDB1 PHF6
ARR
LMO2 Peak Intensity
DU.528
LMO2 Peak Intensity
HSB2
LMO2 Peak Intensity
CCRF-CEM
LMO2 Peak Intensity
0 40
0 5
0 5
0 5

Figure S6


DU.528_PHF6
DU_GATA2
ARR_LDB1
ARR_PHF6
DU.528_PHF6
CCRF_LDB1
CCRF_PHF6
HSB2_PHF6
DU.528_LDB1
CCRF_GATA2
HSB2_GATA2
HSB2_LDB1
ARR_LMO2
ARR_TAL1
CCRF_LMO2
CCRF_TAL1
HSB2_LMO2
HSB2_TAL1
DU.528_LMO2
DU.528_TAL1
DU.528_TAL1
DU.528_LMO2
HSB2_TAL1
HSB2_LMO2
CCRF_TAL1
CCRF_LMO2
ARR_TAL1
ARR_LMO2
HSB2_LDB1
HSB2_GATA2
CCRF_GATA2
DU.528_LDB1
HSB2_PHF6
CCRF_PHF6
CCRF_LDB1
ARR_GATA2
ARR_PHF6
ARR_LDB1
DU.528_GATA2
DU.528_PHF6
1.0
0.8
0.6
0.4
0.2
0.0