BIOINFORMATION
Discovery at the interface of physical and biological sciences

open access www.bioinformation.net Hypothesis

Volume 10(4)

In silico finding of Putative Cis-Acting Elements for the Tethering of Polycomb Repressive Complex2 in Human Genome

Mohammadreza Hajjari[1, 2] *, Mehrdad Behmanesh[2] & Mohammad Mehdi Jahani

[1]Department of Genetics, Shahid Chamran University of Ahvaz, Ahvaz, Iran; [2]Department of Genetics, School of Biological Sciences, Tarbiat Modares University, Tehran, Iran; Mohammadreza Hajjari – Email: Mohamad.hajari@gmail.com; *Corresponding author

Received March 18, 2014; Accepted March 21, 2014; Published April 23, 2014

Abstract:
Polycomb Repressive Complex2 maintains a predetermined state of transcription which constitutes a cellular memory stable over many cell divisions. Since this complex acts through the regulation of chromatin structure, it is important to understand how it is recruited to chromatin. The specific target sequences of this complex such as PRE (polycomb repressive element) have not been completely recognized in human genome. In this study, we have compared the target sequences of this complex with non-target genes in tumor cell lines. Through *in silico* and statistical analyses, we have identified some motifs which are over-represented in target genes against non-target genes. Analyzing these motifs shows some transcription factors which are potential recruiters of Polycomb repressive complex2.

Key words: PRC2, SUZ12, Transcription factors, Epigenetics, *In silico* Analysis.

Background:
Polycomb group (PcG) proteins have long been known to be part of the cellular memory of the cell **[1, 2]**. These proteins are transcriptional repressors which form well defined multimeric complexes and sustain the genes in an off state. Members of PcG proteins are highly conserved and contain histone methyl transferase activity on core histones **[3, 4]**.

Polycomb Repressive Complexes (PRCs) 2/3/4 include histone methyltransferase Enhancer of Zeste protein-2 (EZH2), the Extra Sex Combs protein (EED), the Suppressor of Zeste-12 protein (SUZ12), as well as the histone-binding proteins RbAP46 and RbAP48 **[5-7]**. These proteins are expressed at high levels in embryonic tissues and are required for proper development. Mice lacking these components die during initial development **[8-10]**. However, in normal adult tissues, the expression of SUZ12, EZH2, and EED is low **[11]**. Interestingly, it has been shown that these proteins are present at high levels in a variety of human tumors **[12, 13]**. It seems that these components may be implicated in conferring the neoplastic phenotype **[14]**. Although the misexpression of several Polycomb group proteins in different cancers has been reported, the role of these proteins in cancer development is still poorly understood. Thus, understanding of the way they function will provide important insights into the mechanisms of cancer initiation and progression.

An important question in PcG function revolves around the multiple mechanisms required for appropriate targeting of PRC2 complex in human genome. Hundreds of target genes have been identified in *D. melanogaster* **[15]** and mammalian cells **[11, 16, 17]** by genome wide ChIP-assay and other means. Studies to disclose how PRC1 and PRC2 are recruited to target genes have focused on defining DNA sequence elements, called Polycomb Response Elements (PREs), and transacting factors that identify PREs. PREs have been recognized mainly from the fly *Hox* genes based on their ability to give PcG silencing on reporter genes **[18].** The PRE sequences are poorly conserved in mammals, specifically in the context of *Hox* genes, as analyzed by the PRE prediction program PREdictor **[18].** Due to functional complexity of PRC2 and its interacting proteins, a strict and common target motif in different human cells and tissues has not been introduced yet. Some studies have been done to introduce some motifs similar to PREs. Woo *et al.* found a region named $D_{11.12}$ in hESC (Human Embryonic Stem Cell) which can target PRC2 between HoxD11 and HoxD12 **[19].** Also, Cuddapah *et al.* selected and tested three H3K27me3 enriched regions near three genes in human T cells. They found that these regions can target PcG proteins **[20].** Besides, Cabianca *et al.* introduced D4Z4 elements as sequences sharing several features with PREs **[21].**

Using a variety of ChIP-chip approaches, Squazzo *et al.* have marked a large set of SUZ12 target genes in different human cell lines. They found that SUZ12 target promoters are cell type specific, with transcription factors and homeobox genes dominating in embryonic cells as well as glycoproteins and immunoglobulin-related proteins predominating in adult tumors. They found some target genes common to human cancer cell lines **[17].** In current study, through *in silico* analyses, we focused on these common targets in cancer cell lines to find whether there is any overrepresentation of specific motifs in SUZ12-bound DNA sequences or not. These putative motifs may exert influence on the regulation of genes which have been considered as targets of PRC2. We might get better understanding of the mechanisms of PRC function by analyzing these motifs.

Methodology:

According to the study of Squazzo *et al.* **[17],** sequences of 50 target genes of SUZ12 (from -5kb to Transcription start site) were downloaded from NCBI database. Since Squazzo *et al.* detected SUZ12 in a region between -5kb to transcription start sites of target genes, we used this distance for all of the genes studied. Additionally, the list of 20 housekeeping genes and 40 tissue specific genes were selected randomly from a study by Eisenberg *et al.* **[22]** and "Verygene" database respectively. Similar to target genes of SUZ12, 5Kb upstream regions of housekeeping and tissue specific genes were downloaded from NCBI. Tissue specific and housekeeping genes were considered as the control sets which are not targets of SUZ12. Repeatmasker program (http://www.repeatmasker.org/cgi-bin/WEBRepeatMasker Version: open-3.0.8) was employed to investigate all the repeat classes except simple/low complexity repeats. MEME program available at "http:/ /meme.sdsc.edu/ meme/website/meme-download.html" was used to search for motifs in downloaded sequences. Both one occurrence per sequence (OOPS) and zero or one occurrence per sequence (ZOOPS) models given in MEME were utilized for obtaining motifs. Further, MAST program (http://meme.sdsc.edu/ meme/website/meme-download.html) was used on the motif weight matrix obtained from the MEME program to search for the motifs in the control genes. An e-value cut-off of 0.01 was applied for MAST. Fisher's exact tests were used to test the association of the motifs within upstream sequences of SUZ12 targets against control sets (housekeeping and tissue specific genes). Distinguished motifs were analyzed for the presence of known transcription factor binding site using PATCH program (http://www.gene-regulation.com /pub/programs.html).

Results:

In order to find any motif overrepresentation between upstream sequences of target genes of SUZ12 in cancer cell lines, 50 genes were randomly selected from the study of Squazzo *et al.* These genes are common targets of SUZ12 in testis, breast, and colon cancer cell lines. As a control set, some housekeeping and tissue specific genes for the mentioned tissues were selected. The 5kb upstream sequences of their transcription start sites were repeat-masked for exclusion of complex repeats. Some sequences did not have any complex repeat sequences. In this step, simple repeats were not masked because some transacting factors could be targeted through DNA binding factors that recognize simple sequences **[23, 24].** In order to find motifs in regulatory sequences of 50 genes, the masked sequences were taken as an input for MEME-Chip Program. There were three motifs identified by this analysis. These motifs are shown in **Table 1 (see supplementary material)**.

Motifs 1, 2, and 3 were repeated in 19, 10, and 8 SUZ12 target genes respectively. The motifs identified in targets of SUZ12 were submitted to MAST program to make comparison of them with randomly selected housekeeping and tissue specific genes. The frequency of occurrence for each motif in SUZ12 targets and non-targets sequences with e-value cut off 0.01 is in **Table 2 (see supplementary material).**

All of the motifs showed significant differences between SUZ12 targets and the control genes **(Table 2).** Furthermore, the analysis of the motifs for known transcription factors indicated the presence of binding sites for transcription factors such as Sp1 and GAGA factor.

Discussion:

Attempts to reveal how PRC1 and PRC2 are recruited to target genes have focused on defining DNA sequence elements called Polycomb response elements (PREs), and transacting factors that recognize PREs. The best characterized PREs have been confined to several hundred base pairs, but there is no simple consensus sequence that can provide PRE function. Despite high homology of PcG proteins among different organisms **[25]** and many identified target genes of the PcG proteins in human **[11, 17]**, human PREs have remained elusive until recently **[19, 20].** This can be attributed to the low conservation of recruiters of PRCs such as GAF, PSQ, and Zeste. Identification of more potential PREs in mammals will allow for functional analysis to explore whether transcription factor-mediated recruitment ensures PcG-PRE interaction in mammals.

The rules for targeting PcG proteins to their targets seem likely to vary among different cell types and contexts. Different studies have introduced several sequences which can be targets of PRC2 in different cells. Some of these sequences include D4Z4 and D11.12 **[19-21].** Because Failures in PcG function have intensive effects on cancer progression **[26],** disclosing the role of PcG proteins and the mechanisms they serve in cancer cells is really essential. In this study, we have tried to discriminate the motifs which may potentially tether PRC2 complex onto its common targets in some human cancer cell lines.

Squazzo *et al.* have identified a large set of SUZ12 target genes in different human cell lines. They found some genes that are common targets of SUZ12 in cancer cell lines **[17].** In this study, we found three motifs that are significantly overrepresented in the upstream sequences of the SUZ12 targets as against the non-targets of SUZ12 (including housekeeping and tissue specific genes) in specific tissues. The sequences were identified after masking the repeat sequences by Repeatmasker program. The significant overrepresentation of the motifs in the SUZ12 target genes may propose functional importance of the motifs considering that the motifs identified here are drawn from noncoding sequences including simple repeats but excluding complex repeats through Repeatmasker. It is of note that our study was just restricted to cancer cell lines and some common target genes of PRC2 in them. Target genes of SUZ12 vary in different cell types and SUZ12 may be tethered through some other motifs or mechanisms in other cell lines. But regardless of different potential mechanisms and motifs which target PRC2, we can introduce the motifs identified in this study as potential recruiters of PRC2 complex in some human cancer cell lines.

We observed the occurrence of GAGA binding sites within one-identified motifs (Motif 2) by PATCH program. GAGA factor appears to be a multipurpose transacting factor which is a well known transcription factor **[27]** and has also been shown to be a component of certain types of PcG complexes. In addition, the GAGA factor is recognized as a member of trithorax family of proteins **[28].** The G(A)n motif, GAAAA, which is identified in motif 1 in our study, is recognized as a part of the PRE/TREs sequence for the engagement of PcG complex through DSP1 protein **[29].** The Sp1/KLF binding sites, which are important for PRE activity in *Drosophila*, were also present in motifs 2 and 3. Sp1, which is also a known gene specific transcription factor, has been shown to interact with PRE of *Engrailed* gene. Sp1 has also contributed to the regulation of several genes in breast cancer cells associated with cell growth and cycle progression (cyclin D1, E2F1, c-fos-), angiogenesis (VEGF), and anti-apoptosis (bcl-2) **[30].** According to our analysis, some functions of this factor may be potentially attributed to the tethering of SUZ12 to its targets. In a study by Bengani *et al.*, GAGA factor and Sp1 factor have been introduced as potential transcription factors which may tether SUZ12 protein into its targets in embryonic cells **[31].** Recognizing the binding potential of these transcription factors in identified motifs in our study supports the putative role of these motifs for the tethering of PRC2.

In summary, we predict that the identified motifs in our study could be potentially the sites of interaction of chromatin modifying complexes for epigenetic regulation. Our data could serve as a resource for experimental analysis of binding sites and transacting regulatory complexes interacting with these sites. To our knowledge, this study was the first to find some common sequences as motifs which may potentially target PRC2 in human genome of some cancer cell lines.

Conflict of interest: None

References:

[1] Francis NJ & Kingston RE, *Nat Rev Mol Cell Biol*. 2001 **2**: 409 [PMID: 11389465]
[2] Simon JA & Tamkun JW, *Curr Opin Genet Dev*. 2002 **12**: 210 [PMID: 11893495]
[3] Beisel C *et al. Nature* 2002 **419**: 857 [PMID: 12397363]
[4] Cao R *et al. Science* 2002 **298**: 1039 [PMID: 12351676]
[5] Kuzmichev A *et al. Genes Dev* 2002 **16**: 2893 [PMID: 12435631]
[6] Kuzmichev A *et al. Proc Natl Acad Sci U S A*. 2005 **102**: 1859 [PMID: 15684044]
[7] Cao R & Zhang Y, *Curr Opin Genet Dev* 2004 **14**: 155 [PMID: 15196462]
[8] Pasini D *et al. EMBO J*. 2004 **23:** 4061 [PMID: 15385962]
[9] O'Carroll D *et al. Mol Cell Biol*. 2001 **21**: 4330 [PMID: 11390661]
[10] Faust C *et al. Development* 1995 **121**: 273 [PMID: 7768172]
[11] Kirmizis A *et al. Genes Dev* 2004 **18**: 1592 [PMID: 15231737]
[12] Rajasekhar VK & Begemann M, *Stem Cells*. 2007 **25**: 2498 [PMID: 17600113]
[13] Tsang DP & Cheng AS, *J Gastroenterol Hepatol*. 2011 **26:** 19 [PMID: 21175789]
[14] Varambally S *et al. Nature* 2002 **419:** 624 [PMID: 12374981]
[15] Schuettengruber B & Cavalli G, *Development* 2009 **136**: 3531 [PMID: 19820181]
[16] Bracken AP *et al. EMBOJ*. 2003 **22**: 5323 [PMID: 14532106]
[17] Squazzo SL *et al. Genome Res*. 2006 **16**: 890 [PMID: 16751344]
[18] Ringrose L & Paro R, *Annu Rev Genet*. 2004 **38**: 413 [PMID: 15568982]
[19] Woo CJ *et al. Cell* 2010 **140**: 99 [PMID: 20085705]
[20] Cuddapah S *et al. PLoS One* 2012 **7:** e36365 [PMID: 22570707]
[21] Cabianca DS *et al. Cell* 2012 **149**: 819 [PMID: 22541069]
[22] Eisenberg E & Levanon EY, *Trends Genet*. 2003 **19:** 362 [PMID: 12850439]
[23] Wang YH & Griffith JD, *Proc Natl Acad Sci U S A*. 1996 **93:** 8863 [PMID: 8799118]
[24] Cao H *et al. J Mol Biol*. 1998 **281:** 253 [PMID: 9698546]
[25] Whitcomb SJ *et al. Trends Genet* 2007 **23:** 494 [PMID: 17825942]
[26] Su Y *et al. Epigenetics* 2011 **6:** 16 [PMID: 20818160]
[27] Croston G *et al. Science* **1991** 251: 643 [PMID: 1899487]
[28] Farkas G *et al. Nature* 1994 **371**: 806 [PMID: 7935842]
[29] Dejardin J *et al. Nature* 2005 **434**: 533 [PMID: 15791260]
[30] Safe S & Abdelrahim M, *Eur J Cancer*. 2005 **41:** 2438 [PMID: 16209919]
[31] Bengani H *et al. J Exp Zool B Mol Dev Evol*. 2007 **308:** 384 [PMID: 17358016]

Edited by P Kangueane
Citation: **Hajjari *et al.*** Bioinformation 10(4): 187-190 (2014)

Supplementary material:

Table 1: Motifs identified in SUZ12 targets by MEME Program

Motif	Length	Consensus sequence
1	11	GAAAAGGGAAG
2	30	GGGAGGGGGAGAGGGAGGGGGAGAGTGGGC
3	14	CCACCCCAACACAT

Table 2: Frequency of Motifs in Targets and Non targets genes of SUZ12

Motif	Targets(50)	Non-targets(HKGs*)	Non-targets(TSGs*)	P-value
1	19	0	0	0.0001
2	10	2	1	0.005
3	8	0	0	0.0005

*HKGs: House Keeping Genes (20)
TSGs: Tissue Specific Genes (40)