Data in Brief 13 (2017) 202–207

Contents lists available at ScienceDirect

Data in Brief

journal homepage: www.elsevier.com/locate/dib

Data Article

Datasets on the genomic positions of the MLL1 morphemes, the ZFP57 binding site, and ZFBS-Morph overlaps in the build mm9 of the mouse genome

Minou Bina [a,*], Phillip Wyss [a], Xiaohui C. Song [b]

[a] Purdue University, Department of Chemistry, West Lafayette, IN 47907, USA
[b] Information Technology at Purdue University, West Lafayette, IN 47907, USA

ARTICLE INFO

Article history:
Received 2 May 2017
Received in revised form
19 May 2017
Accepted 23 May 2017
Available online 28 May 2017

Keywords:
CpG-rich motifs
Gene regulation
Genomic imprinting
KMT2A
MLL1 morphemes
Mouse genome
ZFP57 binding site

ABSTRACT

While MLL1 activates gene expression in most tissues, ZFP57 represses transcription. MLL1 selectively interacts with a group of nonmethylated DNA sequences known as the MLL1 morphemes. ZFP57 associates with a methylated hexamer (ZFBS), dispersed in the genomic DNA segments known as Imprinted Control Regions (ICRs) and germline Differentially Methylated Regions (gDMRs), to maintain allele-specific gene repression. We have identified a set of composite DNA elements (ZFBS-Morph overlaps) that provides the sequence context of ZFBS in the canonical ICRs/gDMRs. This report provides tables listing the nucleotide sequences of the MLL1 morphemes and ZFBS-Morph overlaps. The report also offers links to the data repository at Purdue University, for downloading the positions of the MLL1 morphemes, the ZFP57 binding site, and the ZFBS-Morph overlaps in the mouse genome.

DOI of original article: http://dx.doi.org/10.1016/j.ygeno.2017.04.008
* Corresponding author.
E-mail address: bina@purdue.edu (M. Bina).

http://dx.doi.org/10.1016/j.dib.2017.05.050

Specifications Table

Subject area	Genomics
More specific subject area	Gene regulation
Type of data	Tables and text files (in bed format, for display at the UCSC genome browser)
How data was acquired	Analyzing the mouse chromosomes using Perl Scripts
Data format	Tables and text files
Experimental features	None
Data accessibility	Two links to files deposited at the Purdue University Research Repository: 1) Bina, M., Wyss, P.J., Wang, D., Song, X.C. (2014). Localization of MLL1 morphemes in mouse mm9 genomic DNA. Purdue University Research Repository. doi:10.4231/R7KW5CXF https://purr.purdue.edu/publications/1648/1 2) Bina, M., Wyss, P.J., Wang, D., Song, X.C. (2017). Positions of ZFBS and ZFBS-Morph overlaps in the build mm9 of the mouse genome. doi: 10.4231/R7C82782 https://purr.purdue.edu/publications/2473/1

Value of the data

- Two tables and three datasets are offered to the scientific community.
- One table lists the nucleotide sequences of the MLL1 morphemes, the other the nucleotide sequences of ZFBS-Morph overlaps.
- Three datasets were created to provide the genomic positions of functionally important DNA sequence-motifs: the MLL1 morphemes, the ZFP57 binding site, and ZFBS-Morph overlaps.
- The datasets consist of two bed files that could be uploaded onto the UCSC genome browser (build mm9 of the mouse genome), to create custom tracks. One file contains the genomic positions of the MLL1 morphemes, the other includes the genomic positions of ZFP57 binding site and ZFBS-Morph overlaps.
- Availability of these datasets facilitates viewing and analyzing genomic positions of functionally important sequence-motifs in the context of the ENCODE data and mapped landmarks including the position of protein-coding genes and CpG Islands.

1. Data

Mixed Lineage Leukemia 1 (MLL or MLL1) is an essential regulator of transcription [1,2]. MLL1 selectively interacts with a group of nonmethylated DNA sequences known as the MLL1 morphemes: the smallest 'words' in DNA that selectively bind the MT-domain in MLL1 [3]. The *MLL1* gene is one of the mammalian orthologs of the Drosophila *Trithorax* [4]. In human cells, functions of MLL1 include gene bookmarking during mitosis, in a manner favoring genes that were highly transcribed during interphase [5]. Gene bookmarking may involve interactions of MLL1 with morphemes that are localized in CGIs: the CpG islands [3]. The MLL1 morphemes contain 2–3 CpGs and occur in both the forward and the reverse orientation in genomic DNA (Table 1). Even though the MLL1 morphemes are dispersed along the chromosomal DNA, often they are clustered in CGIs [3,6]. Examples include two CGIs (CpG36 and CpG72) associated with the *Plagl1/Zac1 loci* (Fig. 1). As a consequence of length-variability of CGIs [7], morpheme-frequencies in the islands vary: for examples, see Refs. [3,6].