Genomics Data 7 (2016) 264–268

Contents lists available at ScienceDirect

Genomics Data

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

Identification of Tox chromatin binding properties and downstream targets by DamID-Seq

António Miguel de Jesus Domingues [b,1,2], Benedetta Artegiani [a,1,3], Andreas Dahl [b], Federico Calegari [a,*]

[a] DFG-Research Center for Regenerative Therapies, Cluster of Excellence, Faculty of Medicine, TU-Dresden, Germany
[b] Deep Sequencing Group–SFB655, Biotechnology Center, TU-Dresden, Germany

ARTICLE INFO

Article history:
Received 29 January 2016
Received in revised form 1 February 2016
Accepted 1 February 2016
Available online 3 February 2016

Keywords:
DamID-Seq
Neural stem cells
SICER

ABSTRACT

In recent years, DNA adenine methyltransferase identification (DamID) has emerged as a powerful tool to profile protein-DNA interaction on a genome-wide scale. While DamID has been primarily combined with microarray analyses, which limits the spatial resolution and full potential of this technique, our group was the first to combine DamID with sequencing (DamID-Seq) for characterizing the binding loci and properties of a transcription factor (Tox) (sequencing data available at NCBI's Gene Expression Omnibus under the accession number GSE64240). Our approach was based on the combination and optimization of several bioinformatics tools that are here described in detail. Analysis of Tox proximity to transcriptional start sites, profiling on enhancers and binding motif has allowed us to identify this transcription factor as an important new regulator of neural stem cells differentiation and newborn neurons maturation during mouse cortical development. Here we provide a valuable resource to study the role of Tox as a novel key determinant of mammalian somatic stem cells during development of the nervous and lymphatic system, in which this factor is known to be active, and describe a useful pipeline to perform DamID-Seq analyses for any other transcription factor.

© 2016 Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Specifications	
Organism/cell line/tissue	Human Embryonic Kidney 293 (HEK-293 T) cells and mouse neuroblastoma (Neuro-2a) cells
Sex	*not applicable*
Sequencer or array type	Illumina HiSeq 2500
Data format	*Raw and analyzed*
Experimental factors	Fusion Dam-Tox vs. Dam alone
Experimental features	DamID-Seq of the HMG-box transcription factor protein Tox
Consent	*not applicable*
Sample source location	*not applicable*

1. Direct link to deposited data

Data were deposited in Gene Expression Omnibus (GEO) datasets under reference number GSE64240.

http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE64240.

2. Experimental design

During embryonic development of the mammalian cortex, neuroepithelial stem cells expand and generate neurogenic progenitors that in turn divide to give rise to neurons [18]. In an attempt to identify genes involved in controlling this process, our group has generated a double reporter mouse line to isolate the three cell populations present in the developing mouse brain and, by transcriptome analyses, determined their molecular signature [3,4]. The transcription factor Tox was identified among those transcripts that were named "off-switches" since, during corticogenesis, revealed to be highly expressed in neural stem cells, transiently downregulated in neurogenic progenitors, and reinduced in neurons. Many genes essential during neural development are indeed switch genes, showing differential expression between neurogenic progenitors as compared to both neural stem cells and neurons [3]. Interestingly, Tox shared a similar pattern of expression with several such master regulators of corticogenesis. In addition, Tox has been previously associated with differentiation of T-cells in the lymphatic system [1]. Yet, no function was ever reported for this transcription factor during corticogenesis, which leads us to further investigate its role(s) during mammalian brain development.

Acute manipulations of Tox expression in the mouse developing cortex revealed its multiple functions during corticogenesis in different cell types. In fact, Tox negatively regulated neurogenesis by inhibiting differentiation of neural stem cells while, at the same time, controlled

* Corresponding author.
E-mail address: federico.calegari@crt-dresden.de (F. Calegari).
[1] equal contribution.
[2] Current address: Institute of Molecular Biology (IMB), Mainz, Germany.
[3] Current address: Hubrecht Institute, Utrecht, the Netherlands.

http://dx.doi.org/10.1016/j.gdata.2016.02.003
2213-5960/© 2016 Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

neuronal specification and neurites outgrowth of newborn neurons [6]. In order to gain insight into the possible molecular mechanisms by which Tox performs its functions, we then sought to identify its downstream targets by determining binding sites on the genome.

The main experimental approaches to investigate chromatin binding profiles on large scale are chromatin immunoprecipitation (ChIP) and DNA adenine methyltransferase identification (DamID). ChIP has been for many years the gold standard for protein–chromatin interaction profiling [15]. Briefly, ChIP relies on the crosslinking of DNA-protein complexes followed by immunoprecipitation using an antibody recognizing the protein of interest and identification of the co-precipitated DNA sequences. As main drawback, the key determinant for a good ChIP experiment is therefore the availability of an antibody with high affinity and specificity. As alternative antibody-independent method, DamID exploits methylation to label sequences of the genome that are bound by a specific protein. In such a case, the protein of interest is fused to the prokaryotic Dam methylase and expressed in eukaryotic cells at extremely low levels as not to saturate methylation and increase specificity. Dam methylates adenines, modification normally not occurring in eukaryotes, that can therefore be recognized by a restriction enzyme, DpnI, cutting its specific recognition sequence GATC only when the adenine is methylated [16,20,22,23].

Commercially available Tox antibodies were not validated for ChIP and we could not achieve immunoprecipitation with any of the antibodies tested. Therefore, we decided to employ DamID to profile Tox genome binding sites.

Detailed reviews on the advantages and disadvantages of DamID versus ChIP have been reported elsewhere (e.g. [8]). It is however important to point out that DamID typically provided a lower spatial resolution than ChIP-Seq in defining binding domains [14]. Inspired by a report on nuclear envelop proteins combining DamID with high-throughput sequencing [24], we explored the use of DamID-Seq to detect narrow regions of chromatin binding by a transcription factor [6]. Our approach allowed the fine characterization of the binding sites, including binding motif prediction, whose pipeline is provided below.

Experiments were performed according to the standard DamID protocol [23] and using HIV lentiviruses that trigger the expression of Tox-Dam or Dam alone at very low concentrations. Ensuring a low expression level of the ectopic genes is particularly critical to obtain a high signal/background ratio by avoiding saturation of methylation by Dam. In order to achieve this, expression of Tox-Dam fusion gene, and Dam as negative control, was put under the regulation of two inducible promoters neither of which was induced, resulting only in minimal "double-leakyness" of transgene expression. As such, ectopic proteins were undetectable by both immunofluorescence and Western blot (data not shown).

DamID was followed by Illumina next-generation sequencing and bionformatic comparison of Tox-Dam versus Dam alone to identify Tox binding targets. This was performed on both HEK-293T (human embryonic kidney cell line) and Neuro-2a (mouse neuroblastoma cell line). The rationale behind using both cell lines was that in the former case we wanted to use a brain-unbiased system to assess all possible Tox targets while the latter provided a cell line recapitulating the physiology of neural stem cells in which Tox may have additional properties.

Biological duplicates performed on HEK-293T cells revealed to be successful as we could observe substantial differences between Dam-Tox and Dam control samples. This allowed us to identify ca. 13,000 chromatin regions bound by Tox. Conversely, by using the Neuro-2a cell line we did not observe any substantial enrichment of Tox binding as compared to the Dam background. Although we did not perform experiments to explain this discrepancy between cell lines, we find it reasonable that in Neuro-2a the ectopic Tox-Dam fusion protein would compete with the endogenous Tox for its binding to chromatin and since the former is expressed at minimal levels it would be outcompeted by the latter resulting in no differential methylation pattern relative to Dam control [6]. In support to this hypothesis, in HEK-293T cells endogenous Tox expression was undetectable by transcriptome analysis [17] and, hence, ectopic Tox-Dam would be free to bind its targets. These may be important considerations to keep in mind while choosing the cell system to perform DamID, as it could influence the final outcome of the experiments. Tox binding *loci*, including proximity to transcriptional start sites, profiling on enhancers and binding motif were subsequently analyzed by several bioinformatic tools with some of these predictions being later validated in vivo [6].

3. Materials and methods

3.1. Lentiviral transfection of DamID construct

Total RNA was extracted from E13.5 mouse lateral cortices and used as template for RT-PCR amplification of Tox cDNA that was cloned in the pLgwV5EcoDam (Dam construct) [23] to generate the Tox-Dam construct. To ensure that the fusion protein (Tox-Dam) would not display aberrant expression pattern or localization as compared to the native Tox, plasmid coding for the Tox-Dam or the Tox wt protein were transfected in HEK-293T cells and their subcellular localization determined by Western blot and immunocytochemistry [6].

3.2. DamID-Seq

pLgwV5Eco-ToxDam and pLgxV5Dam were used to produce Tox-Dam and Dam control lentiviruses, respectively, as described in [5,7]. HEK-293T and Neuro-2a cells were infected with Tox-Dam or Dam viral supernatant diluted 1:2 or 1:10 and DamID performed as described [23]. Briefly, 48 h after infection genomic DNA was extracted, digested with DpnI, ligated to adaptors and PCR amplified. Experiments were performed in duplicates for each cell line and condition. Sequencing libraries were prepared according to a standard Illumina protocol and subjected to 75 bp single read sequencing on a HiSeq 2000 machine, resulting in ca. 20 million reads per sample (DNA libraries of replicates were sequenced separately).

3.3. Primary processing of sequencing data

Raw read quality was evaluated using FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Read alignment to the genome (*Homo sapiens* ensembl67 or *Mus musculus* ensembl61) was performed with bowtie v0.12.7 [10] using the default parameters with the options "–best" and "-m 1" to retain only uniquely mapped reads. Replicate reproducibility was tested using bamCorrelate from the NGS analysis suite deepTools [13], with the custom options bamCorrelate bins –fragmentLength 200 –corMethod pearson. Pair of replicates displayed Pearson correlation coefficients > 0.80 and therefore the alignments corresponding to 2 replicates of the same condition were merged before peak calling. Since experiments performed in Neuro-2a cells did not show any significant difference between conditions (Tox-Dam vs. Dam), our further analyses were based only on data obtained from HEK-293T cells.

3.4. Identification of Tox binding sites

To identify genomic regions of Tox-Dam enrichment we used the peak caller SICER [27] version 1.1, which has been previously used to detect enrichment in DamID experiments [24]. SICER has been initially developed to detect enrichment (ChIP over input) of diffuse histone modifications. Differently from transcription factors, which usually bind at very localized genomic loci and therefore lead to strong and localized signals, histone modification signals are more diffused and lack well defined peaks. SICER is therefore an algorithm designed to deal with more diffused enrichment spreads over extended genomic regions, rather than strong local enrichment [27]. Since DamID-methylation could spread for some distance from the actual binding site (ca. 2 kb) [20], we thought that

Download English Version:

https://daneshyari.com/en/article/2822189

Download Persian Version:

https://daneshyari.com/article/2822189

Daneshyari.com