



## Data in Brief

## Genome wide mapping of UBF binding-sites in mouse and human cell lines

Jeannine Diesch<sup>a,b</sup>, Ross D. Hannan<sup>a,b,c,d,e,f</sup>, Elaine Sanij<sup>a,b,g,\*</sup><sup>a</sup> Division of Cancer Research, Peter MacCallum Cancer Centre, St. Andrews Place, East Melbourne, Victoria 3002, Australia<sup>b</sup> Sir Peter MacCallum Department of Oncology, University of Melbourne, Parkville, Victoria 3010, Australia<sup>c</sup> Department of Biochemistry and Molecular Biology, Monash University, Clayton, Victoria 3800, Australia<sup>d</sup> Department of Biochemistry and Molecular Biology, University of Melbourne, Parkville, Victoria 3010, Australia<sup>e</sup> Division of Cancer Medicine, Peter MacCallum Cancer Centre, St. Andrews Place, East Melbourne, Victoria 3002, Australia<sup>f</sup> School of Biomedical Sciences, University of Queensland, Brisbane, Queensland 4072, Australia<sup>g</sup> Department of Pathology, University of Melbourne, Parkville, Victoria 3010, Australia

## ARTICLE INFO

## Article history:

Received 16 December 2014

Accepted 29 December 2014

Available online 6 January 2015

## Keywords:

UBF

RNA polymerase I

Histone modifications

ChIP-seq

## ABSTRACT

The upstream binding transcription factor (UBTF, also called UBF) is thought to function exclusively in RNA polymerase I (Pol I)-specific transcription of the ribosomal genes. We recently reported in Sanij et al. (2014) [1] that the two isoforms of UBF (UBF1/2) are enriched at Pol II-transcribed genes throughout the mouse and human genomes. By using chromatin immunoprecipitation coupled with deep sequencing (ChIP-seq) of UBF1/2, Pol I, Pol II, H3K9me3, H3K4me4, H3K9ac and H4 hyperacetylation, we reported a correlation of UBF1/2 binding with enrichments in Pol II and markers of active chromatin. In addition, we examined a functional role for UBF1/2 in mediating Pol II transcription by performing expression array analysis in control and UBF1/2 depleted NIH3T3 cells. Our data demonstrate that UBF1/2 bind highly active Pol II-transcribed genes and mediate their expression without recruiting Pol I. Furthermore, we reported ChIP-seq analysis of UBF1/2 in immortalized human epithelial cells and their isogenically matched transformed counterparts. Here we report the experimental design and the description of the ChIP-seq and microarray expression datasets uploaded to NCBI Sequence Research Archive (SRA) and Gene Expression Omnibus (GEO).

© 2015 The Authors. 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	Murine/NIH3T3; <i>Homo sapiens</i> /mammary epithelial cell lines HMEC and HMLER
Sex	Male or female
Sequencer or array type	Illumina Genome Analyzer II and HiSeq 1000; Affymetrix Mouse Exon ST 1.0 arrays
Data format	Raw ChIP-seq data: FASTQ files; Analyzed ChIP-seq: TXT data; Microarray expression data: CEL files
Experimental factors	The use of antibodies for ChIP-seq and short interfering RNA oligos for knocking down UBF
Experimental features	ChIP-seq and microarray expression
Consent	NA
Sample source location	NA

## Direct link to deposited data

Raw ChIP-seq data is available through the NCBI Sequence Read Archive (SRA, <http://www.ncbi.nlm.nih.gov/sra/>), study accession number (SRP039369). Processed ChIP-seq data and microarray expression data are available through the Gene Expression Omnibus (GEO, <http://www.ncbi.nlm.nih.gov/geo/>), study accession number (GSE63255) and (GSE55461), respectively.

## Experimental design, materials and methods

## Cell culture and RNA interference (RNAi) experiments

Mouse NIH3T3 cells (post-crisis, immortalized embryonic fibroblasts, ATCC) were cultured in Dulbecco's modified Eagle's medium (DMEM) with 10% fetal bovine serum at 37 °C. The human primary mammary epithelial cell line (HMEC) immortalized by expressing TERT, the catalytic subunit of telomerase and the tumorigenic HMLER cell line, an isogenic HMEC-derived cell line expressing the SV40 large-T, TERT, and an

\* Corresponding author at: Growth Control Laboratory, Research Division, Peter MacCallum Cancer Centre, St Andrews Place, East Melbourne, Victoria 3002, Australia. Tel.: +61 3 9656 3758; fax: +61 3 9656 3738.

E-mail address: [elaine.sanij@petermac.org](mailto:elaine.sanij@petermac.org) (E. Sanij).

oncogenic allele of the *HRAS* gene (expressing *HRAS*<sup>V12G</sup>) [2] were cultured in HuMEC ready medium (12752010, life technologies). Dharmafect 2 reagent (Dharmacon) was used to transfect NIH3T3 cells with siRNA at 40 nM according to the manufacturer's protocol. RNA was extracted 48 h after transfection using Qiagen RNA extraction kit. The short interfering RNA (siRNA) oligonucleotide RNA sequences are reported in Sanij et al.[1].

#### Antibodies

Anti-H3K9me3 (ab8898), anti-H3K4me3 (ab8580), and anti-RNA Pol II [4H8] (ab5408) antibodies were obtained from Abcam. Anti-hyperacetylated H4 (06-946) and anti H3K9ac (07-352) antibodies were from Upstate (Millipore). Antibodies targeting UBF1/2 and the largest subunit of the Pol I complex (POLR1A/RPA194) were raised in-house and were used as reported in Sanij et al.[3].

#### Chromatin immunoprecipitation

ChIP was carried out as described [4]. Cross-linking was achieved with 0.6% formaldehyde for 10 min at 37 °C. The reaction was stopped by adding 0.125 M glycine and cells were collected and washed with PBS. Cell pellets were resuspended in 10 mM Tris pH 7.4, 10 mM NaCl, 10 mM MgCl<sub>2</sub>, 0.5% NP-40 and protease inhibitor cocktail and incubated on ice for 10 min. The suspensions were centrifuged and pellets were resuspended in SDS lysis buffer (50 mM Tris pH 8.1, 10 mM EDTA, 1% SDS and protease inhibitor cocktail) at  $4 \times 10^6$  cells per 300  $\mu$ l. Sonication was performed using Covaris (Covaris Inc) for 25 min at (Duty cycle 20%, intensity 5, cycles per burst 200; 30 s ON, 30 s OFF) to obtain chromatin shearing range between 200–400 base pairs.

Sonicated chromatin was centrifuged at 13,000 rpm for 10 min at 4 °C and supernatants were collected for immunoprecipitation (IP). 150  $\mu$ l of lysates corresponding to  $2 \times 10^6$  cells per IP was diluted to 1.5 ml with IP dilution buffer (1 mM DTT, 0.01% SDS, 1.1% Triton X-100, 1.2 mM EDTA, 16.7 mM Tris pH 8.1, 167 mM NaCl and protease inhibitor cocktail) and 120  $\mu$ l was used as reference (8% of input genomic DNA (gDNA)). The lysates were precleared with 35  $\mu$ l of protein A agarose/salmon sperm DNA (Upstate, Millipore). For all ChIPs, 4  $\mu$ g of purified antibody or 8  $\mu$ l of sera was used per IP and incubated overnight at 4 °C with rotation. 50  $\mu$ l of protein A agarose/salmon sperm DNA beads was added per tube and incubated for 2 h at 4 °C with rotation. Beads were washed sequentially with three different wash buffers and two washed in TE buffer for 5 min at 4 °C with rotation (low salt wash buffer: 20 mM Tris pH 8.1, 2 mM EDTA, 1% Triton X-100, 0.1% SDS, 150 mM NaCl; high salt wash buffer: 20 mM Tris pH 8.1, 2 mM EDTA, 1% Triton X-100, 0.1% SDS, 500 mM NaCl; LiCl wash buffer: 10 mM Tris pH 8, 1 mM EDTA, 0.25 M LiCl, 0.5% NP-40, 0.5% Deoxycholate (sodium salt); TE buffer: 10 mM Tris pH 8, 1 mM EDTA). The immunoprecipitated chromatin was eluted twice with 250  $\mu$ l elution buffer (1% SDS, 0.1 M NaHCO<sub>3</sub>) with rotation for 15 min at room temperature. 20  $\mu$ l of 5 M NaCl was added and samples were incubated at 65 °C overnight to reverse protein–DNA crosslinking. Protein K digestion was then performed and DNA was extracted using (1:1) phenol:chloroform extraction followed by isopropanol precipitation. The % of immunoprecipitated DNA was calculated relative to the reference gDNA input. The quality of enrichments in binding was measured relative to rabbit sera ChIP controls.

DNA concentration was measured with the Qubit dsDNA HS assay kit (Invitrogen).

#### Sequencing experiments

Sequencing libraries of 10–30 ng of ChIPed DNA and input gDNA were prepared using the TruSeq ChIP sample preparation kit as per manufacturer's protocol (Illumina, San Diego, CA, USA). NIH3T3 ChIP-seq libraries (UBF1/2, Pol I, Pol II, H3K4me3, H3K9me3, H3K9ac,

H4 hyperacetylation, gDNA) and HMEC ChIP-seq libraries (UBF1/2, Pol I, gDNA) were sequenced using the Illumina Genome Analyzer II platform at the Peter MacCallum Cancer Centre, while HMLER ChIP-seq libraries of UBF1/2 and gDNA were performed using Illumina HiSeq 1000.

Base calling was performed using CASAVA-1.8.2 (Illumina) with default parameters and sequencing reads mapped to the mouse mm9 or human hg19 genome assembly using the Burrows–Wheeler Aligner (BWA) [5]. After removing duplicate reads with Picard tools, peaks over input gDNA were called using MACS1.4 (Model-based Analysis of ChIP-seq) [6]. Normalized fold change was calculated for each peak and summit using the Bioconductor R package [7] and peaks were annotated to RefSeq genes using the R package ChIPpeakAnno [8].

#### Quality control and functional annotation

Only peak regions with a FDR below 10% and p-value below 0.00001 were selected, and significant peaks were validated by qPCR as recently reported in Sanij et al.[1]. The Bioconductor R package and the Sole-Search program [11] were used to determine the distribution of UBF1/2 peaks relative to RefSeq genes and their transcription start and termination sites (TSS and TTS).

#### Expression analysis

We performed microarray analysis (Affymetrix, Mouse Exon ST 1.0 arrays) on three biological replicates of UBF1/2 knockdown samples using two independent siRNA oligos, a non-silencing sirEGFP or Mock transfected NIH3T3 samples. The arrays were normalized using Robust Multi-Array Average expression measure (RMA) [9] and differential expression was then determined using a linear model and the Limma package (<http://bioinf.wehi.edu.au/limma/>). Moderated t-statistics were generated and significance was assessed using log fold change and an FDR (false discovery rate) adjusted p-value [10].

Correlation between expression values from the mock NIH3T3 samples for all genes, genes with significant UBF1/2 peaks <2 kb from their TSS, or genes with no UBF1/2 binding at their TSS, was performed using the R package and statistical significance assessed using the t-tests.

## Results

Genome-wide enrichment of UBF1/2 showed preference for binding near TSSs, whereas no significant binding preference was observed at the TTSs. Almost 40% of all UBF1/2 binding overlapped with first exons and introns of annotated genes in mouse as well as human cells.

We intersected regions bound by UBF1/2 with a variety of posttranslational histone modification binding sites as well as Pol II and Pol I enriched regions in NIH3T3 cells. Overlapping peaks were defined as peaks with at least 1 bp overlap. While little correlation between UBF1/2 binding and the presence of the transcriptional repressive mark H3K9me3 was observed, almost 50% of the UBF1/2 peaks overlapped with the activating H3K4me3, H4K9ac and H4 hyperacetylation marks and Pol II enrichment. Thus, UBF1/2 is preferably bound to open chromatin structures associated with active promoters and gene bodies. Furthermore, the comparison of UBF1/2 and Pol I ChIP-seq analysis in NIH3T3 and HMEC cell lines revealed little overlap in binding (less than 8% and 3%, respectively), demonstrating that UBF1/2 does not recruit Pol I to Pol II genes.

We then investigated whether UBF1/2 binding correlated with transcriptional activity by intersecting UBF1/2-bound genes with gene expression data in NIH3T3 cells. This revealed that genes enriched in UBF1/2 are expressed at high levels compared to non-UBF1/2 bound genes and all transcribed genes in the genome. Gene ontology analysis using the Genomic Regions Enrichment of Annotation Tool (GREAT) [11] revealed chromatin assembly and nucleosome organization and assembly as the biological processes most significantly enriched with

UBF1/2-bound genes and common between the mouse and human datasets.

To determine if UBF1/2 binding has a functional effect on gene expression, we intersected the ChIP-seq and microarray expression datasets and identified genes whose expression was significantly altered by UBF1/2 knockdown and were bound by UBF1/2 within 500 bp of their TSSs. Gene ontology analysis using the MetaCore pathways software (Thomson Reuters) identified a significant overrepresentation of genes belonging to chromatin/nucleosome assembly and DNA packaging, including canonical histone genes and histone gene variants indicating that their transcription may be directly regulated by UBF1/2. In Sanij et al.[1], we further validated a novel role for UBF1/2 in mediating Pol II transcription of histone genes.

## Discussion

We recently reported a dual function for the Pol I transcription factor UBF1/2 in the regulation of Pol I and Pol II mediated transcription [1]. In addition, we described a fundamental role for UBF1/2 in regulating highly transcribed Pol II genes, including the histone gene clusters. Moreover, in the transformed HMLER cells, we demonstrated that UBF1/2 is enriched at an additional cohort of genes involved in DNA damage and repair including mediators of ATR/ATM-regulated DNA damage response, signal transduction by TP53 and G1 to S transition of cell cycle. Thus, UBF1/2 binding is dynamic, context dependent and potentially associated with malignant transformation.

In summary, we demonstrated a fundamental role for UBF1/2 in coupling Pol I transcription and the cell's capacity to grow with the fidelity of chromatin assembly through its ability to coordinately regulate the expression of some of the most highly transcribed Pol I and Pol II genes in the genome including the histone clusters and ribosomal DNA repeats.

## Conflict of interest statement

We declare no conflict of interest.

## References

- [1] E. Sanij, J. Diesch, A. Lesmana, G. Poortinga, G. Lidgerwood, N. Hein, D.P. Cameron, J. Ellul, G.J. Goodall, L.H. Wong, et al., A novel role for the Pol I transcription factor UBF1 in maintaining genome stability through the regulation of highly transcribed Pol II genes. *Genome Res.* Dec 1. pii: gr.176115.114. [Epub ahead of print].
- [2] B. Elenbaas, L. Spirio, F. Koerner, M.D. Fleming, D.B. Zimonjic, J.L. Donaher, N.C. Popescu, W.C. Hahn, R.A. Weinberg, Human breast cancer cells generated by oncogenic transformation of primary mammary epithelial cells. *Genes Dev.* 15 (2001) 50–65.
- [3] E. Sanij, G. Poortinga, K. Sharkey, S. Hung, T.P. Holloway, J. Quin, E. Robb, L.H. Wong, W.G. Thomas, V. Stefanovsky, et al., UBF levels determine the number of active ribosomal RNA genes in mammals. *J. Cell Biol.* 183 (2008) 1259–1274.
- [4] G. Poortinga, K.M. Hannan, H. Snelling, C.R. Walkley, A. Jenkins, K. Sharkey, M. Wall, Y. Brandenburger, M. Palatsides, R.B. Pearson, et al., MAD1 and c-MYC regulate UBF and rDNA transcription during granulocyte differentiation. *EMBO J.* 23 (2004) 3325–3335.
- [5] H. Li, R. Durbin, Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* 25 (2009) 1754–1760.
- [6] Y. Zhang, T. Liu, C.A. Meyer, J. Eeckhoutte, D.S. Johnson, B.E. Bernstein, C. Nussbaum, R.M. Myers, M. Brown, W. Li, X.S. Liu, Model-based Analysis of ChIP-Seq (MACS). *Genome Biol.* 9 (2008).
- [7] R.C. Gentleman, V.J. Carey, D.M. Bates, B. Bolstad, M. Dettling, S. Dudoit, B. Ellis, L. Gautier, Y. Ge, J. Gentry, et al., Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol.* 5 (2004) R80.
- [8] L.J. Zhu, H. pages, C. Gazin, N. Lawson, S. Lin, D. Lapointe, M. Green, ChIPpeakAnno: Batch Annotation of the Peaks Identified From Either ChIP-seq or ChIP-Chip Experiments. R Package Version 1.6.0. 2009.
- [9] R.A. Irizarry, B.M. Bolstad, F. Collin, L.M. Cope, B. Hobbs, T.P. Speed, Summaries of affymetrix GeneChip probe level data. *Nucleic Acids Res.* 31 (2003).
- [10] G. Smyth, Limma: linear models for microarray data. in: V.C. Gentleman, S. Dudoit, R. Irizarry, W. Huber (Eds.), *Bioinformatics and Computational Biology Solutions using R and Bioconductor*, R, Springer, New York, 2005, pp. 397–420.
- [11] C.Y. McLean, D. Bristor, M. Hiller, S.L. Clarke, B.T. Schaar, C.B. Lowe, A.M. Wenger, G. Bejerano, GREAT improves functional interpretation of cis-regulatory regions. *Nat. Biotechnol.* 28 (2010) 495–501.