Genomics Data 4 (2015) 1–7

Contents lists available at ScienceDirect

Genomics Data

journal homepage: http://www.journals.elsevier.com/genomics-data/

Data in Brief

Dynamics of GATA1 binding and expression response in a GATA1-induced erythroid differentiation system

Deepti Jain [a,b,1], Tejaswini Mishra [a,b,1], Belinda M. Giardine [a,b], Cheryl A. Keller [a,b], Christopher S. Morrissey [a,b], Susan Magargee [a,b], Christine M. Dorman [a,b], Maria Long [a,b], Mitchell J. Weiss [c], Ross C. Hardison [a,b,*]

[a] Center for Comparative Genomics and Bioinformatics, Pennsylvania State University, University Park, PA 16802, USA
[b] Department of Biochemistry and Molecular Biology, Pennsylvania State University, University Park, PA 16802, USA
[c] Department of Hematology, St Jude Children's Research Hospital, Memphis TN 38105, USA

ARTICLE INFO

Article history:
Received 16 January 2015
Accepted 18 January 2015
Available online 29 January 2015

Keywords:
Gene regulation
ChIP-seq
Transcription factor
DNA binding
RNA-seq

ABSTRACT

During the maturation phase of mammalian erythroid differentiation, highly proliferative cells committed to the erythroid lineage undergo dramatic changes in morphology and function to produce circulating, enucleated erythrocytes. These changes are caused by equally dramatic alterations in gene expression, which in turn are driven by changes in the abundance and binding patterns of transcription factors such as GATA1. We have studied the dynamics of GATA1 binding by ChIP-seq and the global expression responses by RNA-seq in a GATA1-dependent mouse cell line model for erythroid maturation, in both cases examining seven progressive stages during differentiation. Analyses of these data should provide insights both into mechanisms of regulation (early versus late targets) and the consequences in cell physiology (e.g., distinctive categories of genes regulated at progressive stages of differentiation). The data are deposited in the Gene Expression Omnibus, series GSE36029, GSE40522, GSE49847, and GSE51338.

Specifications	
Organism/cell line	*Mus musculus* cell lines G1E and G1E-ER4
Strain	129
Sex	Male
Sequencer or array type	Illumina Genome Analyzer IIx, Illumina HiSeq 2000
Data format	Sequence reads: fastq; mapped reads: bam, bai (bam index file); peaks calls: broadPeak; signal tracks: bigwig
Experimental factors	Mouse cell line (G1E) with a genetic knockout of the *Gata1* gene and a daughter cell line (G1E-ER4) transduced with an estrogen-activated *Gata1*-estrogen receptor transgene that can rescue erythroid maturation in an estrogen-dependent manner.
Experimental features	G1E cells (untreated) and G1E-ER4 cells treated with 10 nM estradiol for six time points (0, 3, 7, 14, 24, and 30 h) were (a) analyzed genome-wide for binding by GATA1 using ChIP-seq (including control samples of input DNA at each time point) and (b) mapped for transcription genome-wide by strand-specific, paired-end RNA-seq of poly A+ RNA, including biological replicates.
Consent	Not applicable
Sample source location	Not applicable

Direct link to deposited data

All data are available through ENCODE data portals:
https://www.encodeproject.org
http://www.mouseencode.org

The GATA1-ChIP-seq data sets are available in three GEO Series: GSE51338, GSE36029, and GSE49847.
http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE51338
http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE36029
http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE49847

The RNA-seq data sets are available in three GEO Series: GSE40522, GSE51338, and GSE49847.
http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE40522
http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE51338
http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE49847

The individual data sets and links are listed in Table 1.

Experimental design, materials, and methods

Cell lines used

G1E cells are an immortalized *Gata1* null cell line derived from embryonic stem cells [1], and the daughter cell line G1E-ER4 has been stably rescued by transduction with a virus expressing a hybrid gene

* Corresponding author at: 304 Wartik Laboratory, Penn State University, University Park, PA 16802, USA.
E-mail address: rch8@psu.edu (R.C. Hardison).
[1] These authors contributed equally.

http://dx.doi.org/10.1016/j.gdata.2015.01.008

Table 1
Genomic data sets and URLs for access.

Cell	Treatment with estradiol	Feature	Replicates*	GEO accession URL
G1E	Untreated	Paired-end RNA-seq	2	http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM995536
G1E-ER4	Untreated	Paired-end RNA-seq	2	http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM995532
G1E-ER4	3 h	Paired-end RNA-seq	2	http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM995538
G1E-ER4	7 h	Paired-end RNA-seq	2	http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM995531
G1E-ER4	14 h	Paired-end RNA-seq	2	http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM995527
G1E-ER4	24 h	Paired-end RNA-seq	2	http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM995539
G1E-ER4	30 h	Paired-end RNA-seq	2	http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM995541
G1E	Untreated	GATA1 ChIP-seq	2	http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM923581
G1E-ER4	Untreated	GATA1 ChIP-seq	1	http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM995445
G1E-ER4	3 h	GATA1 ChIP-seq	1	http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM995443
G1E-ER4	7 h	GATA1 ChIP-seq	1	http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM995442
G1E-ER4	14 h	GATA1 ChIP-seq	1	http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM995444
G1E-ER4	24 h	GATA1 ChIP-seq	1	http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM923572
G1E-ER4	30 h	GATA1 ChIP-seq	1	http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM995448
G1E	Untreated	Input seq	2	http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM923580
G1E-ER4	Untreated	Input seq	1	http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM995441
G1E-ER4	3 h	Input seq	1	http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM995437
G1E-ER4	7 h	Input seq	1	http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM995436
G1E-ER4	14 h	Input seq	1	http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM995440
G1E-ER4	24 h	Input seq	1	http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM995439
G1E-ER4	30 h	Input seq	1	http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM995438

* Number.

encoding the GATA1-ER protein [2,3]. Both G1E and untreated G1E-ER4 cells proliferate and show many properties of immature erythroid progenitor cells [2,4]. Upon treatment with an estrogen such as estradiol (E2), G1E-ER4 cells mature synchronously and rapidly, recapitulating many aspects of normal erythroid differentation in a manner dependent on activation of GATA1-ER [2,4–6]. Among the changes during differentation are a loss of proliferative capacity, a reduction in cell size, condensation of the nucleus, increase and then decrease in CD44, and an increase in TER119 [7] (Fig. 1, *left*).

Cell culture methods

G1E and G1E-ER4 cells were grown in IMDM media with 15% fetal calf serum 2 U/ml erythropoietin (EpoGen from Amgen) and 50 ng/ml stem cell factor [1,2]. To induce erythroid maturation, G1E-ER4 cells were treated with 10^{-8} mol/L β-estradiol for 3, 7, 14, 24, and 30 h. Cells were harvested by centrifugation at 500×*g* for 5 min at 4 °C and washed once in 1× PBS.

Chromatin immunoprecipitation (ChIP)

ChIP assay was performed as previously described [2]. Briefly, 75 million cells in 1× PBS were cross-linked for 10 min by adding formaldehyde at a final concentration of 0.4%, and glycine was added at a final concentration of 125 mM to quench cross-linking. Cells were then lysed (10 mM Tris–HCl, pH 8.0, 10 nM NaCl, 0.2% NP40) for 10 min on ice, washed once in 1× PBS, followed by nuclear lysis (50 mM Tris–HCl 8.0, 10 mM EDTA, 1% SDS) for 10 min on ice. Chromatin was then diluted further with Immunoprecipitation Buffer (20 mM Tris–HCl, pH 8.0, 2 mM EDTA, 150 mM NaCl, 1% Triton X-100, 0.01% SDS) and a 1× Protease Inhibitor Cocktail set V, EDTA-free (Calibiochem, La Jolla, CA). A Misonix S-4000 sonicator was used to shear samples in 8 repeats of 30 cycles of 1 s on, 1 s off sonication at 30% output power 30 on ice. Fragments in the size range of 200-400 base pairs were obtained. Sonicated chromatin was pre-cleared overnight at 4 °C with 20 μg rat non-immune sera (IgG) on protein G agarose beads. Ten micrograms of the rat anti-GATA1 (sc-265, Santa Cruz Biotechnology, Santa Cruz, CA; lot number L1609) antibody were also pre-bound to protein G agarose beads overnight at 4 °C. For binding, pre-cleared chromatin was added to the antibody–bead complex and incubated with rotation at 4 °C for 4 h; 200 μL of pre-cleared chromatin was saved for use as input. After binding, the beads were washed with Wash Buffer I (20 mM Tris–HCl, pH 8.0, 2 mM EDTA, 50 mM NaCl, 1% Triton X-100, 0.1% SDS), High-Salt Wash Buffer (20 mM Tris–HCl, pH 8.0, 2 mM EDTA, 500 mM NaCl, 1% Triton X-100, 0.1% SDS), Wash Buffer II (10 mM Tris–HCl, pH 8.0, 1 mM EDTA, 250 mM LiCl, 1% NP40, 1% deoxycholate), and 1× TE. DNA:protein complexes were then eluted from beads with Elution Buffer (1% SDS, 100 mM NaHCO3). Reverse cross-linking of immunoprecipitated chromatin was accomplished by the addition of NaCl to ChIP and input samples, followed by incubation overnight at 65 °C with 1 μg RNase A. To remove proteins, each sample was treated with 6 μg Proteinase K for 2 h at 45 °C. Immunoprecipitated DNA was finally purified using the Qiagen PCR Purification Kit.

Illumina library preparation for ChIP-Seq

All samples including input were processed for library construction for Illumina sequencing using Illumina's ChIP-seq Sample Preparation Kit. In brief, DNA fragments were repaired to generate blunt ends, and a single 'A' nucleotide was added to each end. Double-stranded Illumina adaptors were ligated to the fragments. Ligation products were amplified by 18 cycles of PCR, and the DNA between 250 and 350 base pairs was gel purified. Completed libraries were quantified with Quant-iT dsDNA HS Assay Kit. The DNA libraries were sequenced on the Illumina Genome Analyzer IIx or HiSeq 2000 as indicated (Table 2) using Illumina's kits and reagents as appropriate.

Mapping for ChIP-Seq

Raw ChIP-seq reads were first groomed using FASTQ Groomer on Galaxy [8–10].This program verifies that each base call has a corresponding quality value, and that the quality value is in the Sanger, Phred + 33 format. Groomed reads were then mapped to mouse mm9 genome using Bowtie [11] using the parameters –$m = -1$ (no limit), $-k = 1$, $-$y, and $-$best, thus allowing reads to map to multiple locations, but reporting only the single, best alignment. This option was chosen to allow reads to map in duplicated regions.

Peak calling for ChIP-seq

The mapped reads for each time point in G1E-ER4 cells and untreated G1E cells were then passed to MACS [12] with the matched control (input) data set for peak calling using an mfold of 12, *p*-value threshold of 1e−05 and bw (bandwidth) set to 120. We filtered ChIP-seq peaks