Supplemental Bioinformatic Methods

Sequencing of Immunoprecipitated Chromatin

Libraries were generated robotically with 10 ng of fragmented DNA (100-300 bp) using the Kapa HTP Library Preparation Kit (Kapa Biosystems) as per the manufacturer's recommendations except that adapters and PCR primers were diluted 100-fold, the size selection step was done after the PCR step and the number of PCR cycles increased by 6. Adapters and PCR primers were purchased from Integrated DNA Technologies whereas size selection has been performed on a Pippin Prep instrument (SAGE Biosciences Inc). Libraries were quantified using the Quant-iT™ PicoGreen® dsDNA Assay Kit (Life Technologies) and the Kapa Illumina GA with Revised Primers-SYBR Fast Universal kit (D-Mark Biosciences, Toronto ON). Average size fragment was determined using a LaChip GX (PerkinElmer, Waltham MA) instrument. Libraries were sequenced on a single-end 100bp read length run on a HiSeq2000 (Illumina).

Analysis of ChIP-sequencing

Illumina HiSeq2000 .bam files provided by the Genome Quebec were processed using rmdup (version 1.0.0) to remove duplicate reads. Files were converted to FASTQ format using Bam-to-FASTA/Q (version 1.0.0) and reads groomed to remove Illumina adapter sequences with TrimGalore!(1) in Galaxy Suite (2-4). Single-end reads aligned to reference MGSCv37 (MM9) using Bowtie (5). Bowtie reported only the best alignment for each fragment (--best) with a maximum number of 2 mismatches with an average quality score of at least 70. Peaks called in MACS 1.4.1 with an mfold of (10,30), tag size --100, band width --300, genome size --1865500000, p-value 1e-05, wiggle resolution --10 (6). Peak-calling was conducted for two biological replicates for each condition, and replicate region files used for determining differential

peaks between the experimental conditions. Peaks significantly differing between non-induced and two induced samples were determined using the R package DiffBind (7) with a p-value of 0.05 and a fold-change greater than 2. Wiggle files for control and doxycycline-induced were generated from merged tag data from replicate experiments using MACS1.4.1 as described above. Common regions between induced and uninduced conditions were determined using Intersect (Galaxy Tool Version 1.0.0) and reporting the intervals of two datasets that overlapped for at least 100bp. Location of peaks within the genome were annotated using CEAS (version 1.0.0) (8), span:3000, profiling resolution:50, Promoter/downstream lower, middle, and upper intervals of 1000, 3000, and 10000 (respectively). Bi-Promoter lower range: 2500, Bi-Promoter upper range:5000, Relative distance:3000.

Association of differentially-bound regions with gene expression

Regions significantly enriched by ChIP were correlated with gene expression as determined by Affymetrix Mouse Gene 2.0 ST Microarray (9). Differentially expressed genes were defined as those with a 1.5-fold change in expression between conditions and P value $\leq$0.05 based on ANOVA using Partek software suite. Peaks increased or new upon induction were correlated to target genes using GREAT (10), using the Basal plus extension rule with the proximal region was adjusted to 15kb upstream. Genes were considered differentially regulated by PU.1 if they demonstrated significant changes by both microarray and upstream binding of PU.1. Genes differentially expressed and associated with altered PU.1-binding were annotated using DAVID (11), and top biological processes reported from GOTERM_BP_FAT. Interactions between these differentially expressed genes was visualised using String-DB *(Search Tool for the Retrieval of Interacting Genes/Proteins)* using all gene symbols and organism: mus musculus.

Determination of sequence motifs

Genomic sequences from significantly enriched ChIP regions were extracted using Extract Genomic DNA (Galaxy Tool Version 2.2.3) using locally cached MM9. FASTA files were sorted in order of confidence (best to worst). Motifs were identified using Meme-suite (meme-suite.org), using uniprobe_mouse input motifs, expecting any number of repetitions and reporting 5 motifs. Frequency of primary motifs was determined using SPAMO, reporting the number of sequences containing the primary motif as a percentage of all sequences submitted.

Sequencing of small RNAs

Total RNA was quantified using a NanoDrop Spectrophotometer ND-1000 (NanoDrop Technologies, Inc.) and its integrity was assessed using a 2100 Bioanalyzer (Agilent Technologies). Libraries were generated from 1000 ng of total RNA using the TruSeq® Small RNA Sample Prep Kit (Illumina), as per the manufacturer's recommendations. cDNA construct purification was performed on a Pippin Prep instrument (SAGE Science, Beverly MA). Final libraries were quantified using the Quant-iT™ PicoGreen® dsDNA Assay Kit (Life Technologies) and the Kapa Illumina GA with Revised Primers-SYBR Fast Universal kit (D-Mark Biosciences, Toronto ON). Average fragment size was determined using a LaChip GX (PerkinElmer, Waltham MA) instrument. Libraries were multiplexed and sequenced on an Illumina HiSeq 2000 SR50 using the services of Genome Quebec.

Analysis of small RNA sequencing

Illumina TRUSeq RNA sequencing adapters were removed from sequencing reads using TrimGalore! (1). Trimmed reads were processed and mapped to the mouse genome (MM10) using MirDeep2 Mapper (V2.0.0) discarding reads less than 15 bp. Collapsed reads were mapped to

mirBase Release 21 (12) using MirDeep2 for identification of novel and known microRNAs (13). Known miRBase mouse microRNAs were quantified using MiRDeep2 quantifier (V2.0.0) and differential analysis of fold change was conducted using DESeq2 (14). Differentially expressed microRNAs were considered significant if they had a log-fold change greater than 1 and a p-value <0.05.

Analysis of predicted miRNA targets

Significantly over-expressed miRNAs determined using DeSeq were analysed for predicted and validated targets using MirGate (15). MirGate database was queried using MirBase miRNA identifiers against Organism: mouse. Targets were considered validated if they had a least 1 confirmed prediction. Computational Predictions of 3' UTRs were determined using TargetScan and Miranda.

1. **Krueger F.** 2015. Trim Galore!
2. **Blankenberg D, Von Kuster G, Coraor N, Ananda G, Lazarus R, Mangan M, Nekrutenko A, Taylor J.** 2010. Galaxy: a web-based genome analysis tool for experimentalists. Curr Protoc Mol Biol **Chapter 19:**Unit 19 10 11-21.
3. **Giardine B, Riemer C, Hardison RC, Burhans R, Elnitski L, Shah P, Zhang Y, Blankenberg D, Albert I, Taylor J, Miller W, Kent WJ, Nekrutenko A.** 2005. Galaxy: a platform for interactive large-scale genome analysis. Genome Res **15:**1451-1455.
4. **Goecks J, Nekrutenko A, Taylor J.** 2010. Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. Genome Biol **11:**R86.

5. **Langmead B, Trapnell C, Pop M, Salzberg SL.** 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol **10:**R25.

6. **Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W, Liu XS.** 2008. Model-based analysis of ChIP-Seq (MACS). Genome Biol **9:**R137.

7. **Stark R, Brown G.** 2011. DiffBind: differential binding analysis of ChIP-Seq peak data. http://bioconductororg/packages/release/bioc/vignettes/DiffBind/inst/doc/DiffBindpdf.

8. **Shin H, Liu T, Manrai AK, Liu XS.** 2009. CEAS: cis-regulatory element annotation system. Bioinformatics **25:**2605-2606.

9. **Ziliotto R, Gruca MR, Podder S, Noel G, Ogle CK, Hess DA, DeKoter RP.** 2014. PU.1 promotes cell cycle exit in the murine myeloid lineage associated with downregulation of E2F1. Experimental Hematology **42:**204-217.e201.

10. **McLean CY, Bristor D, Hiller M, Clarke SL, Schaar BT, Lowe CB, Wenger AM, Bejerano G.** 2010. GREAT improves functional interpretation of cis-regulatory regions. Nat Biotechnol **28:**495-501.

11. **Huang da W, Sherman BT, Lempicki RA.** 2009. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc **4:**44-57.

12. **Kozomara A, Griffiths-Jones S.** 2014. miRBase: annotating high confidence microRNAs using deep sequencing data. Nucleic Acids Res **42:**D68-73.

13. **Friedlander MR, Chen W, Adamidi C, Maaskola J, Einspanier R, Knespel S, Rajewsky N.** 2008. Discovering microRNAs from deep sequencing data using miRDeep. Nat Biotechnol **26:**407-415.

14. **Love MI, Huber W, Anders S.** 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol **15:**550.

15. **Andrés-León E, González Peña D, Gómez-López G, Pisano DG.** 2015. miRGate: a curated database of human, mouse and rat miRNA–mRNA targets. Database: The Journal of Biological Databases and Curation **2015:**bav035.