Genomics Data 5 (2015) 126–128

Contents lists available at ScienceDirect

Genomics Data

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

Data in Brief

Global gene expression changes in liver following hepatocyte nuclear factor 4 alpha deletion in adult mice

Sumedha Gunewardena [a,b], Chad Walesky [c], Udayan Apte [c,*]

[a] Department of Biostatistics, University of Kansas Medical Center, Kansas City, KS, USA
[b] Department of Molecular and Integrative Physiology, University of Kansas Medical Center, Kansas City, KS, USA
[c] Department of Pharmacology, Toxicology and Therapeutics, University of Kansas Medical Center, Kansas City, KS, USA

ARTICLE INFO

Article history:
Received 15 May 2015
Received in revised form 26 May 2015
Accepted 28 May 2015
Available online 5 June 2015

Keywords:
Cell proliferation
Ect2
Gene repression
Liver regeneration
AAV8

ABSTRACT

Hepatocyte nuclear factor 4 alpha (HNF4α) is known as the master regulator of hepatic differentiation, which regulates over 60% of the hepatocyte specific genes. Recent studies including this (Walesky et al. Am J Physiol Gastrointest Liver Physiol. 304:G26-37, 2013) demonstrated that HNF4α also inhibits hepatocyte proliferation via repression of pro-mitogenic genes. In this study hepatocyte specific HNF4α knockout mice were generated using 2–3 month old HNF4α-floxed mice treated with Cre recombinase under Major Urinary Protein promoter delivered in AAV8 vector (MUP-iCre-AAV8). Control mice were treated with MUP-EGFP-AAV8. Livers were isolated from control and KO mice one week after AAV8 administration and used for gene array analysis. These data revealed several new negative target genes of HNF4α, majority of which are pro-mitogeneic genes inhibited by HNF4α in adult hepatocytes.

Specifications	
Organism/cell line/tissue	HNF4α floxed mice (mixed background)
Sex	Male
Sequencer or array type	Affymetrix's GeneChip Mouse Genome 430 2.0 arrays
Data format	CEL files and RMA normalized files
Experimental factors	Wild type (WT) vs. Knockout (KO)
Experimental features	HNF4α was deleted in adult male HNF4α-floxed mice (HNF4α-floxed/floxed) by injecting Cre recombinase under the control of Major Urinary protein (MUP) promoter carried by a AAV8 virus vector (MUP-iCre-AAV8). Control mice were given MUP-EGFP-AAV8. Samples were taken one week after virus injection.
Consent	Level of consent allowed for reuse if applicable (typically for human samples)
Sample source location	Kansas City, KS USA

1. Direct link to deposited data

http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE35782.

2. Experimental design, materials and methods.

2.1. Microarray analysis.

The global differences in genes expression between HNF4α$^{Fl/Fl}$ mice treated with the MUP-iCre-AAV8 vector and those treated with the MUP-EGFP-AAV8 control vector were measured using Affymetrix's GeneChip Mouse Genome 430 2.0 arrays. These arrays consist of over 45,000 probe sets representing over 34,000 well characterized mouse genes. Each probe set consists of 11 pairs of probes with each pair consisting of a perfect matched and mismatched 25-mer oligonucleotide interrogating the 3' end of a transcript. The mismatch probe differs from the perfect match probe by a single base substitution at the center of the probe and is used to determine the level of nonspecific hybridization. The Affymetrix chip definition file (CDF) used in the analysis was Mouse430_2.cdf and the annotation file used was Mouse430_2.na27.annot.csv.

2.2. Array normalization

We used a single knock-out sample (mice treated with MUP-iCre-AAV8 vector) and two control (mice treated with MUP-EGFP-AAV8 control vector) biological replicate samples generated from pooled (100 mg per mouse) livers of three individual mice each for microarray analysis. The microarrays were background-corrected using the robust multi-

* Corresponding author at: Department of Pharmacology, Toxicology and Therapeutics, University of Kansas Medical Center, 3901 Rainbow Blvd, MS1018, Kansas City, KS 66160, USA. Tel.: +1 913 588 9247.
E-mail address: uapte@kumc.edu (U. Apte).

http://dx.doi.org/10.1016/j.gdata.2015.05.037

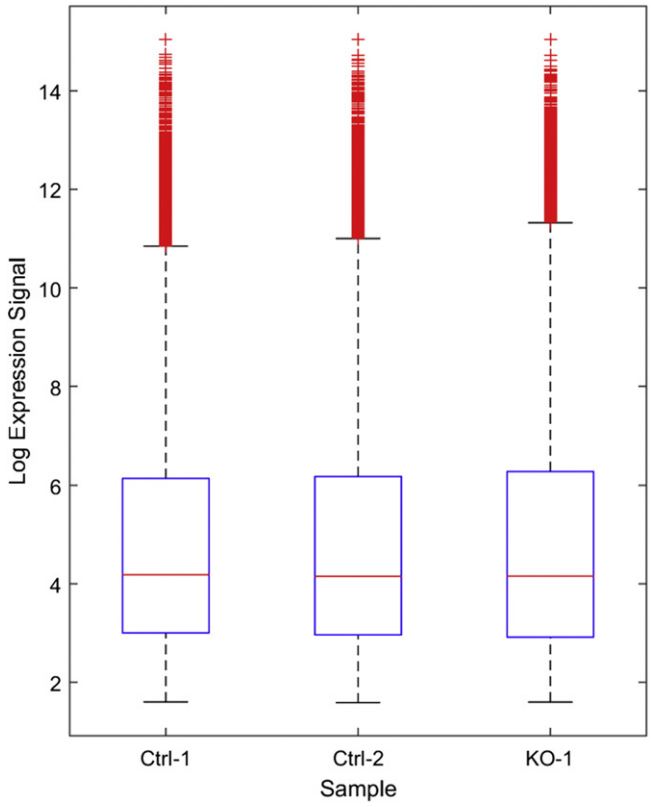

Fig. 1. Boxplot of the background-corrected, normalized and summarized intensity values.

array average (RMA) background correction model [1]. This model assumes that the observed intensity (z) is the sum of the true intensity (x) distributed exponentially, and random noise distributed normally (y), $z = x + y$. The background-corrected intensity values were quantile normalized across all chips, making their probe intensity distributions the same. The resulting intensity values were log (base 2) transformed. Probesets were then summarized using the median polish method [2]. A boxplot of the background-corrected, normalized and summarized intensity values is shown in Fig. 1.

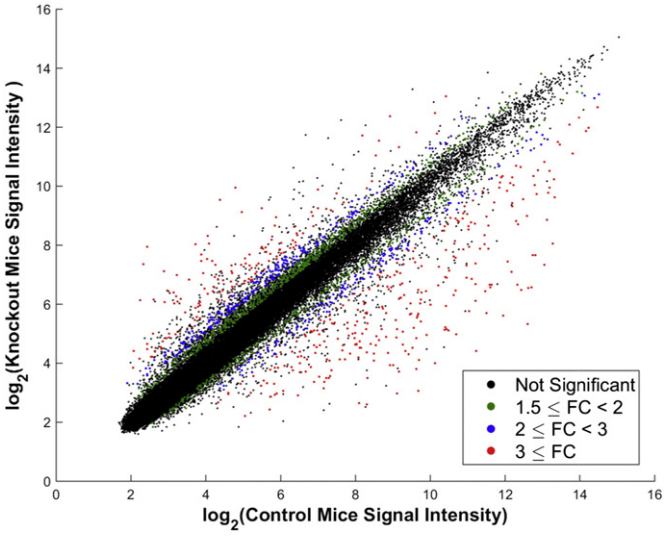

Fig. 3. Scatterplot of significantly (p-value ≤ 0.05) differentially expressed genes.

2.3. Differential expression analysis

The principal components analysis (PCA) plot shown in Fig. 2 demonstrates significant global differences in gene expression between the knock-out and control samples. For each transcript, the log (*base 2*) ratio of the difference in expression between knock-out and control was calculated by subtracting the average log control transcript intensity from the knock-out transcript intensity. This value was transformed to a linear scale by taking two to the power of the value giving the ratio $r = 2^{(K_i - C_i)}$ where K_i is the log signal intensity of the knock-out transcript i and C_i, the average log signal intensity of the control transcript i. This ratio was converted to a fold change using the formula:

$$FC(r) = \begin{cases} -1/r, & r < 0 \\ r, & r \geq 0 \end{cases}$$

The statistical significance of the observed fold change was calculated by fitting a 1-way ANOVA model by using the method of moments [3]. The model is described by the equation $Y_{ij} = \mu + CATEGORY_i + \varepsilon_{ij}$ where Y_{ij} represents the jth observation on the ith category. The

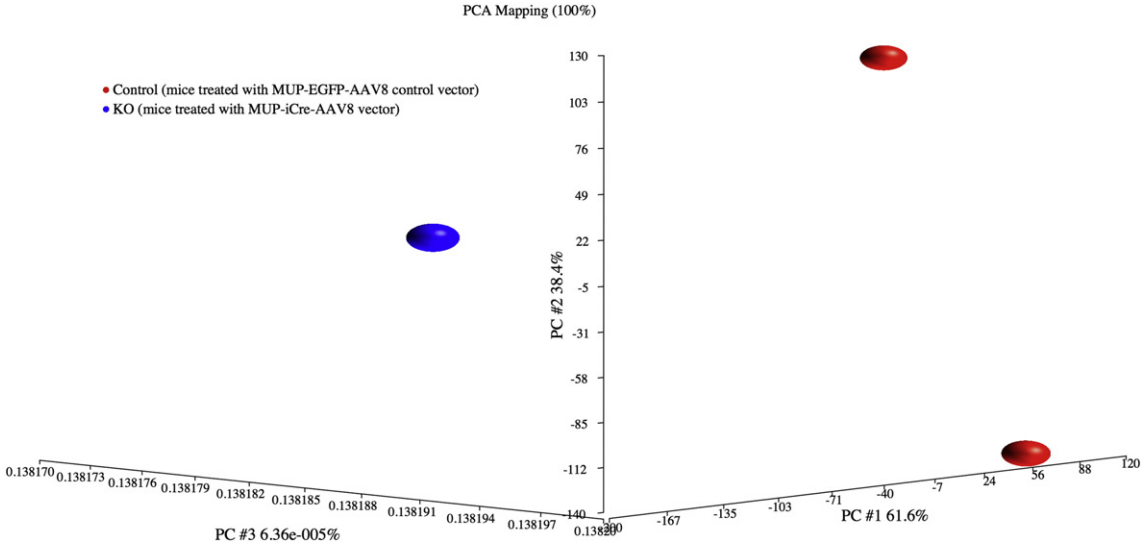

Fig. 2. The principal components analysis (PCA) plot of the control and knockout samples.

Table 1
Mapping statistics.

	Hnf4a ChIP	Input
Read length	36 bp single end	36 bp single end
Total reads	24,011,998	25,108,375
Reads aligned exactly 1 time	8,551,655 (35.61%)	8,901,068 (35.45%)
Reads aligned > 1 time	5,166,444 (21.52%)	6,010,408 (23.94%)
Overall alignment	57.13%	59.39%

intercept, μ, models the common effect for the whole experiment. *CATEGORY* is a categorical variable representing the knock-out and control transcript and ε_{ij} represents the random error present in the jth observation of the ith category. The errors ε_{ij} are assumed to be normally and independently distributed with mean 0 and a fixed standard deviation, ϑ, for all measurements. A scatterplot showing the significantly (p-value ≤ 0.05) differentially expressed genes are shown in Fig. 3.

2.4. Chromatin immunoprecipitation-sequencing data analysis.

In order to distinguish those genes specifically regulated by Hnf4a among the significantly perturbed genes obtained from microarray, we analyzed previously published chromatin immunoprecipitation (ChIP)-sequencing (ChIP-Seq) data for Hnf4a targets obtained from the National Center for Biotechnology Information Short Read Archive study SRA008281 [4]. The following experiments from this study were used in our analysis; Hnf4a: SRX003308, INPUT: SRX020706 and SRX020707. The raw reads were mapped to the mouse reference genome (NCBI37/mm9) using bowtie-0.12.3 [5]. The mapping statistics for the Hnf4a ChIP and Input samples are shown in Table 1. Peak detection was performed using the Model-based Analysis of ChIP-Seq (MACS) algorithm [6] with the peak detection p-value cutoff set at 1e−5 (default). These resulting binding sites were filtered for significant sites based on a false discovery rate cutoff set at 10%. We searched for the Hnf4a consensus sequence within a 250 bp region from either side of the called peaks using a weight-matrix match with at least 80% similarity. The Hnf4a weight matrix was obtained from the JASPAR database [7]. Binding sites were annotated by PeakAnalyzer [8] using the nearest TSS option. Significantly differentially expressed genes with an Hnf4a binding target within 10 kb of its transcriptional start site were identified as putative Hnf4a target genes.

3. Funding

These studies were supported by NIH — P20 RR021940, 5T32E8007079, R01DK098414 and AASLD/ALF Liver Scholar Award (Udayan Apte). The gene array studies were performed by the Kansas University Medical Center-Microarray Facility (KUMC-MF), which is supported by the Kansas University-School of Medicine, KUMC Biotechnology Support Facility, the Smith Intellectual and Developmental Disabilities Research Center (HD02528), and the Kansas IDeA Network of Biomedical Research Excellence (RR016475).

References

[1] R.A. Irizarry, et al., Exploration, normalization, and summaries of high density oligonucleotide array probe level data. Biostatistics 4 (2) (2003) 249–264.
[2] J.W. Tukey, Exploratory data analysis. Addison-Wesley Series in Behavioral Science, Reading, xvi, Addison-Wesley Pub. Co., Mass., 1977 688 p.
[3] C. Eisenhart, The assumptions underlying the analysis of variance. Biometrics 3 (1) (1947) 1–21.
[4] B.G. Hoffman, et al., Locus co-occupancy, nucleosome positioning, and H3K4me1 regulate the functionality of FOXA2-, HNF4A-, and PDX1-bound loci in islets and liver. Genome Res. 20 (8) (2010) 1037–1051.
[5] B. Langmead, et al., Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 10 (3) (2009) R25.
[6] Y. Zhang, et al., Model-based analysis of ChIP-Seq (MACS). Genome Biol. 9 (9) (2008) R137.
[7] A. Sandelin, et al., JASPAR: an open-access database for eukaryotic transcription factor binding profiles. Nucleic Acids Res. 32 (Database issue) (2004) D91–D94.
[8] M. Salmon-Divon, et al., PeakAnalyzer: genome-wide annotation of chromatin binding and modification loci. BMC Bioinf. 11 (2010) 415.