# Integrated Approach for the Identification of Human Hepatocyte Nuclear Factor 4alpha Target Genes Using Protein Binding Microarrays

Eugene Bolotin, Hailing Liao, Tuong Chi Ta, Chuhu Yang, Wendy Hwang-Verslues, Jane R. Evans, Tao Jiang, Frances M. Sladek

Hepatology, 2010

**In this file:**

Supplemental Materials and Methods

**Supplemental Materials and Methods**

***Reagents.*** The expression vector pMT7.rHNF4α2 containing wild type (wt) rat HNF4α2 (NM_022180), the predominant isoform in liver, has been previously described (1), as have the vectors containing the human HNF4α2 (NM_000457) (pcDNA3.1.hHNF4α2) and HNF4α8 (pcDNA3.1.hHNF4α8) (2, 3) and the affinity purified antibody to the very C-terminus of HNF4α (α445) (4). Mouse monoclonal antibodies to the N-terminus (αNTD) (PP-K9218-00) and C-terminus of HNF4α2 (αCTD) (PP-H1415-00) and secondary antibodies conjugated to Cy5 (Northern Lights 637 Fluorochrome-labeled donkey anti-mouse or anti-rabbit, #NL008 or NL005, respectively) were purchased from R&D Systems (Minneapolis, MN).

***Cell culture conditions***. Monkey kidney cells (COS-7, ATCC #CRL-1651) were maintained in Dulbecco's Modified Eagle's Medium (DMEM) supplemented with 10% Bovine Calf Serum (BCS). Human hepatocellular carcinoma/hepatoblastoma cell line (HepG2, ATCC# HB-8065) were maintained in DMEM containing 10% Fetal Bovine Serum (FBS), 1% non-essential amino acids, and 1% sodium pyruvate. All cell lines were supplemented with 1% penicillin/streptomycin and maintained in a 5% $CO_2$ incubator at 37 °C.

***Protein binding microarray (PBM) design and primer extension.*** Custom 8x15k arrays of single-stranded 42- to 51-mer oligonucleotides were manufactured by Agilent Technologies (Santa Clara, CA). Both PBM1 and PBM2 contained 3000 unique sequences replicated five times. Each sequence begins with a 27-mer linker (5'-TCGACCGATACTCTAATCTCCCTAGGC-3') followed by a variable region of 5 to 14

nucleotides and a 5-nt cap (5'-GCGCG-3') (Fig. 1C). PBM1 contained a combination of sequences collected from the literature, mined from several ChIP-chip datasets (5, 6), predicted by previously developed models (7), and created from variations on the consensus 5'-AGGTCAaAGGTCA-3'. Random controls and Sp1 sites were included on the array to account for nonspecific and indirect binding (for a complete list of sequences on PBM1, see Supplemental Table S2A). PBM2 contained sequences derived from PBM1, sequences obtained from SVM1 (see below) searches on human promoter regions, and the ChIP-chip data set (5) (for a complete list of sequences on PBM2, see Supplemental Table S2B). The primer extension reaction was performed using Sequenase 2.0 (USB, Cleveland, OH), dUTP-Cy3 (GE Healthcare, Waukesha, WI), and a common primer (5'-TCGACCGATACTCTAATCTCCC-3') as previously described (8) except for the following modifications: the hybridization chamber was inverted and incubated at 85 °C for 10 min, followed by 75 °C for 10 min, 65 °C for 10 min, 50 °C for 10 min, and 55 °C for 90 min. The chamber was disassembled, washed extensively in PBS (pH 7.5) and air dried at room temperature. The dUTP incorporation was visualized as described below.

***PBM application.*** PBMs were pre-moistened in PBS plus 0.01% Triton-X 100 for 1 min and blocked for 1 hr with PBS plus 2% nonfat dry milk, subsequently washed for 10 min in PBS plus 0.1% Tween-20 and then incubated for 1 hr with protein binding solution (16 mM HEPES, pH 7.8,120 mM KCl, 8 mM EDTA, 8 mM EGTA, 1% Tween-20, 0.25 µg/µl of poly-dIdC) and 500 ng HNF4α in crude nuclear extracts from transfected COS-7 cells (see above). The arrays were washed with PBS plus 0.1% Tween-20 for 5 min at low speed on a shaking platform and incubated with primary antibody (αNTD, αCTD or α445) diluted 1:100 in PBS, 2% milk, 0.1% Tween-20 overnight at room temperature. Following incubation with secondary antibody (GαM

or GαR conjugated to Cy5) at 1:50 in the same buffer for 1 hr, the arrays were washed 3x in PBS plus 0.1% Tween-20 for 5 min, then 3 min in PBS, air dried and scanned using a GenePix Axon 4000B scanner (Molecular Devices, Sunnyvale, CA) at 543 nm (Cy3) dUTP and 633 nm (Cy5 conjugated secondary antibody). All PBS and milk was filtered through 0.45 μm filters (Corning, Lowell, MA); all other reagents were filtered through 0.33 μm filters (M2135, MoBiTec, Gottingen, Germany). All washes and incubations were performed in an Agilent hybridization chamber at room temperature (27 °C).

***PBM analysis and data normalization.*** All PBMs were scanned using a GenePix Axon 4000B scanner (Molecular Devices, Sunnyvale, CA) at 543 nm (Cy3) to check for even primer extension, and 633 nm (Cy5) to quantify protein binding. Scanning was performed with a 5 μm resolution at optimal laser intensity and at near-saturation of the highest intensity spots. Images were saved as lossless TIFF files and quantified using GenePix 6.0 software (Molecular Devices). Aberrant spots were manually flagged and removed from subsequent analysis. Background-subtracted mean intensities were calculated for remaining spots. The signals were gradient-corrected using Micro-Array NORmalizatation of array-CGH data (MANOR) implemented in R (9) as needed. Cross- and intra-array normalization was performed using quantile normalization (10), enabling comparison between independent experiments. Replicates for each probe were averaged and their coefficient of variation (CV) was calculated. Only probes with a CV less than 0.3 were used for the training set.

***SVM training and sequence analysis.*** The training data from PBM1 was generated by averaging six correlated arrays (Supplemental Table S2A and Fig. S3); the sequences were ranked based on

their relative intensity. A kernel-based support vector machine (KSVM) function from Kernlab package in R with Laplace dot kernel was used to train the SVM1 model (11). SVM1 was applied in a sliding window approach to classify the 13-mer sequences in all of the annotated human promoters (UCSC hg18) and the ChIP-chip dataset into "binding" and "nonbinding" categories (5). The threshold was empirically adjusted until the false positive rate (~5%) and false negative rate (~5%) were simultaneously achieved in the 10-fold cross validation test. The top predicted binding sequences with a promoter score >0.41 and a ChIP-chip score >0.25 (resulting in ~1,700 and ~1,500 sequences, respectively) were selected for PBM2 (see Supplemental Table S2B). A second SVM, which is called SVM2 and uses the regression mode, was trained on three averaged PBM2 experiments (Fig. S4) and achieved a high correlation with PBM2 in the 10-fold cross validation test ($R^2 = 0.75$) (Fig. 4D). The SVM2 model was also used to search the promoter regions (-2kb to +1 kb relative to the transcription start site, +1) of all annotated genes in the human genome (UCSC hg18) following a sliding window approach.

***RNA interference and expression profiling analysis.*** RNA interference (RNAi) against HNF4α1 was performed in HepG2 cells using siRNAs corresponding to nucleotides +179 to +197 of human *HNF4A* (NM_178849, sense siRNA: 5'-UGUGCAGGUGUUGACGAUGdTdT-3', antisense siRNA 5'-CAUCGUCAACACCUGCACAdTdT-3') purchased from Dharmacon Research, Inc. (An NCBI Blast search indicated that this sequence is unique to *HNF4A*). Approximately 24 hr prior to transfection, cells were plated at a density of ~1.5 $x10^5$ or ~2.5 $x10^5$ cells/well in a 12- or 6-well plate, respectively (~50-70% confluency) without antibiotics. The siRNAs (100 or 200 pmol, respectively) were introduced into the cells using TransIT-TKO transfection reagent purchased from Mirus Bio Corporation (Madison, WI). Each experiment

included a control containing just the transfection reagent as well as a nonspecific siRNA control against firefly luciferase (PGL3: sense siRNA 5'-CUUACGCUGAGUACUUCGAdTdT-3'; antisense siRNA 5'-dTdTGAAUGCGACUCAUGAAGCU). To verify HNF4α protein levels, cells were lysed in RIPA buffer purchased from Santa Cruz Biotechnology and analyzed by SDS-PAGE followed by immunoblotting (Supplemental Fig. S5). Total RNA was extracted using Trizol Reagent (Invitrogen, Carlsbad, CA) and reverse transcribed using the Reverse Transcription System kit (Promega, Madison, WI). PCR amplification was performed in the linear range using a PTC-100 TM programmable thermal controller (MJ Research, Inc., Hercules, CA). One-fifth of each reaction was electrophoresed on a 2% agarose gel, stained with ethidium bromide and visualized by UV light (see Supplemental Table S3B for a list of PCR primers used). Expression profiling analysis was performed with Affymetrix oligonucleotide arrays (HGU133 Plus 2.0) using RNA from control (PGL3 siRNA) or treated (HNF4α siRNA) HepG2 cells (48 h, 200 pmol siRNA per well of 6-well plate, introduced by Lipofectamine 2000 (Invitrogen)). Arrays were hybridized in biological replicate by the UCR Genomics Core Instrumentation Facility. Results were analyzed by Bioconductor LIMMA package (12).

***Gel shift conditions.*** Electrophoretic mobility shift analysis (EMSA, or gel shift) was carried out for Figures 5D, S6 and S7C essentially as previously described (13) using crude nuclear extracts from COS-7 cells transfected with HNF4α expression vectors and $^{32}$P-radiolabelled probes as indicated. Unlabelled competitor oligonucleotides were added prior to the addition of the HNF4α protein in the indicated amounts; α445 antibody specific to HNF4α was added ~15 min after the HNF4α protein. In Figure S7C, competitor oligonucleotides were prepared by USB Sequenase 2.0 (USB) extension as was done for the PBMs and the shift conditions were modified to mimic

the PBM conditions (most notably addition of Tween-20 to 0.01%). See Supplemental Table S8 for sequences of all oligonucleotides used in gel shift assays.

**References**


1. Yuan X, Ta TC, Lin M, Evans JR, Dong Y, Bolotin E, Sherman MA, et al. Identification of an endogenous ligand bound to a native orphan nuclear receptor. PLoS ONE 2009;4:e5609.
2. Chartier FL, Bossu JP, Laudet V, Fruchart JC, Laine B. Cloning and sequencing of cDNAs encoding the human hepatocyte nuclear factor 4 indicate the presence of two isoforms in human liver. Gene 1994;147:269-272.
3. Eeckhoute J, Formstecher P, Laine B. Maturity-onset diabetes of the young Type 1 (MODY1)-associated mutations R154X and E276Q in hepatocyte nuclear factor 4alpha (HNF4alpha) gene impair recruitment of p300, a key transcriptional co-activator. Mol Endocrinol 2001;15:1200-1210.
4. Sladek FM, Zhong WM, Lai E, Darnell JE, Jr. Liver-enriched transcription factor HNF-4 is a novel member of the steroid hormone receptor superfamily. Genes Dev 1990;4:2353-2365.
5. Odom DT, Dowell RD, Jacobsen ES, Nekludova L, Rolfe PA, Danford TW, Gifford DK, et al. Core transcriptional regulatory circuitry in human hepatocytes. Mol Syst Biol 2006;2:2006 0017.
6. Rada-Iglesias A, Wallerman O, Koch C, Ameur A, Enroth S, Clelland G, Wester K, et al. Binding sites for metabolic disease related transcription factors inferred at base pair resolution by chromatin immunoprecipitation and genomic microarrays. Hum Mol Genet 2005;14:3435-3447.
7. Ellrott K, Yang C, Sladek FM, Jiang T. Identifying transcription factor binding sites through Markov chain optimization. Bioinformatics 2002;18 Suppl 2:S100-109.
8. Bulyk ML. Analysis of sequence specificities of DNA-binding proteins with protein binding microarrays. Methods Enzymol 2006;410:279-299.
9. Neuvial P, Hupe P, Brito I, Liva S, Manie E, Brennetot C, Radvanyi F, et al. Spatial normalization of array-CGH data. BMC Bioinformatics 2006;7:264.
10. Bolstad BM, Irizarry RA, Astrand M, Speed TP. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. Bioinformatics 2003;19:185-193.
11. Karatzoglou A, Smola A, Hornik K, Zeileis A. kernlab – An S4 Package for Kernel Methods in R. Journal of Statistical Software 2004;11.
12. Gentleman R, Carey V, Dudoit S, Irizarry R, Huber W, editors. Limma: linear models for microarray data. New York,: Springer; 2005.
13. Jiang G, Nepomuceno L, Hopkins K, Sladek FM. Exclusive homodimerization of the orphan receptor hepatocyte nuclear factor 4 defines a new subclass of nuclear receptors. Mol Cell Biol 1995;15:5131-5143.