



Published in final edited form as:

*Hepatology*. 2010 February ; 51(2): 642–653. doi:10.1002/hep.23357.

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

Eugene Bolotin<sup>1</sup>, Hailing Liao<sup>4</sup>, Tuong Chi Ta<sup>2</sup>, Chuhu Yang<sup>1,\*</sup>, Wendy Hwang-Verslues<sup>3,†</sup>, Jane R. Evans<sup>4</sup>, Tao Jiang<sup>5,6</sup>, and Frances M. Sladek<sup>4,6</sup>

<sup>1</sup>Genetics, Genomics and Bioinformatics Graduate Program, University of California Riverside, Riverside, CA

<sup>2</sup>Cell, Molecular, and Developmental Biology Graduate Program, University of California Riverside, Riverside, CA

<sup>3</sup>Environmental Toxicology Graduate Program, University of California Riverside, Riverside, CA

<sup>4</sup>Department of Cell Biology and Neuroscience, University of California Riverside, Riverside, CA

<sup>5</sup>Department of Computer Science and Engineering, University of California Riverside, Riverside, CA

<sup>6</sup>Institute for Integrated Genome Biology (also at UCR)

### Abstract

Hepatocyte nuclear factor 4 alpha (HNF4 $\alpha$ ), a member of the nuclear receptor superfamily, is essential for liver function and is linked to several diseases including diabetes, hemophilia, atherosclerosis, and hepatitis. Although many DNA response elements and target genes have been identified for HNF4 $\alpha$  the complete repertoire of binding sites and target genes in the human genome is unknown. Here, we adapt protein binding microarrays (PBMs) to examine the DNA-binding characteristics of two HNF4 $\alpha$  species (rat and human) and isoforms (HNF4 $\alpha$ 2 and HNF4 $\alpha$ 8) in a high-throughput fashion. We identified ~1400 new binding sequences and used this dataset to successfully train a Support Vector Machine (SVM) model that predicts an additional ~10,000 unique HNF4 $\alpha$ -binding sequences; we also identify new rules for HNF4 $\alpha$  DNA binding. We performed expression profiling of an HNF4 $\alpha$  RNA interference knockdown in HepG2 cells and compared the results to a search of the promoters of all human genes with the PBM and SVM models, as well as published genome-wide location analysis. Using this integrated approach, we identified ~240 new direct HNF4 $\alpha$  human target genes, including new functional categories of genes not typically associated with HNF4 $\alpha$ , such as cell cycle, immune function, apoptosis, stress response, and other cancer-related genes.

**Conclusion**—We report the first use of PBMs with a full-length liver-enriched transcription factor and greatly expand the repertoire of HNF4 $\alpha$ -binding sequences and target genes, thereby

---

Copyright © 2009 by the American Association for the Study of Liver Diseases.

Address reprint requests to: Frances M. Sladek, Ph.D., Professor of Cell Biology, Department of Cell Biology and Neuroscience, 2115 Biological Sciences, University of California, Riverside, Riverside, CA 92521-0314. frances.sladek@ucr.edu; fax: 951-827-3087.

\*Current address: Roche Diagnostics Shanghai Limited, 1045 Central Huaihai Road, Shanghai 200031, China.

†Current address: Genomics Research Center, Academia Sinica, No. 128 Academia Road, Section 2, Nankang District, Taipei 115, Taiwan.

Potential conflict of interest: Nothing to report.

Additional Supporting Information may be found in the online version of this article.

identifying new functions for HNF4 $\alpha$ . We also establish a web-based tool, HNF4 Motif Finder, that can be used to identify potential HNF4 $\alpha$ -binding sites in any sequence.

Hepatocyte nuclear factor 4 $\alpha$ , HNF4 $\alpha$  (*HNF4A*), is a member of the nuclear receptor superfamily of ligand-dependent transcription factors (NR2A1) and a liver-enriched transcription factor (TF) that is also expressed in the kidney, pancreas, intestine, colon, and stomach.<sup>1</sup> Originally identified based on its ability to bind DNA response elements in the human apolipoprotein C3 (*APOC3*) and mouse transthyretin (*Ttr*) promoters,<sup>2</sup> HNF4 $\alpha$  has since been shown to play a critical role in both the development of the embryo and the adult liver.<sup>3,4</sup> Mutations in the *HNF4A* coding sequence and promoter regions are linked to Maturity Onset Diabetes of the Young 1 (MODY1),<sup>5</sup> and mutations in HNF4 $\alpha$  response elements have been directly linked to disease, most notably in genes encoding blood coagulation factors in hemophilia and in HNF1 $\alpha$  in MODY3.<sup>6–8</sup> Through classical promoter analysis, functional HNF4 $\alpha$ -binding sites have been identified in >140 genes, including those involved in the metabolism of glucose, lipids, and amino acids, as well as xenobiotics and drugs<sup>1,4,9</sup> (see Supporting Table 1A for a listing of those genes). Recent genome-wide location analyses suggest that the number of HNF4 $\alpha$  targets may be much greater (>1000) based on widespread binding of HNF4 $\alpha$  to promoter regions,<sup>10–12</sup> although it is not known how many of those are functional targets. A more comprehensive list of direct HNF4 $\alpha$  targets was recently made even more critical with our finding that HNF4 $\alpha$  binds an exchangeable ligand and hence may be a potential drug target.<sup>13</sup>

HNF4 $\alpha$  binds DNA exclusively as a homodimer.<sup>14,15</sup> The canonical HNF4 $\alpha$  consensus sequence consists of the half site AGGTCA with one nucleotide spacer (referred to as a DR1, AGGTCA $\times$ AGGTCA).<sup>16</sup> Whereas the number of experimentally verified HNF4 $\alpha$  binding sequences is sizable (>217) (Supporting Tables 1A and 1B), they were derived in a biased fashion building on the first HNF4 $\alpha$ -binding sites,<sup>2</sup> and subsequently on the direct repeat rules for nuclear receptor DNA binding.<sup>16</sup> Furthermore, the total number of 13-base oligomer (13-mer) permutations is much greater than 217 ( $4^{13} \sim 67$  million), and whereas HNF4 $\alpha$  will certainly not bind all potential 13-mers, the total number of DNA sequences that will bind HNF4 $\alpha$  is anticipated to be in the tens of thousands. Because the presence of one or more HNF4 $\alpha$  response elements in the promoter region of a gene is a prerequisite for classification as a direct HNF4 $\alpha$  target, it is desirable to accurately predict all the HNF4 $\alpha$ -binding sites throughout the genome in an unbiased fashion.

Recent genome-wide technologies, most notably genome-wide location analysis (i.e., chromatin immunoprecipitation [ChIP] followed by tiling arrays, known as “ChIP-chip”) and expression profiling, have greatly accelerated the identification of target genes for many TFs, including HNF4 $\alpha$ . However, as powerful as those technologies are, they provide information only about the state of the cells used in the assay, not about any other physiological or pathological state. Furthermore, expression profiling cannot indicate whether a gene is a direct or an indirect target and ChIP does not provide any information about whether the gene is expressed by the bound TF. And neither assay allows one to precisely identify the sequence to which the TF binds. The third tool in the genomic arsenal—computational prediction of target genes—is curiously less developed than the other two. Although many attempts have been made at predicting TF binding sites, including our own for HNF4 $\alpha$ ,<sup>17</sup> this approach still suffers from a lack of sizable datasets of verified binding sites.

To improve the prediction of potential HNF4 $\alpha$  target genes, we adapted the protein binding microarray (PBM) technology to rank thousands of HNF4 $\alpha$  sequences based on their relative binding affinities using full-length protein expressed in mammalian cells. We compare two species of HNF4 $\alpha$  (rat and human) and two tissue-specific isoforms (HNF4 $\alpha$ 2

and HNF4 $\alpha$ 8). Additionally, we use a Support Vector Machine (SVM), a powerful machine learning model to predict additional HNF4 $\alpha$ -binding sequences with high accuracy. Finally, we combine the PBM and SVM binding site searches with expression profiling performed here and CHIP-chip performed by others to identify ~240 new direct target genes of HNF4 $\alpha$  in cells of hepatic origin (see Fig. 1A for an overview).

## Materials and Methods

See Supporting Materials and Methods for additional details.

### Preparation of HNF4 $\alpha$ Proteins in COS-7 Cells

Nuclear extracts were prepared from COS-7 cells transiently transfected with HNF4 $\alpha$  expression vectors as previously described.<sup>15</sup> Mock-transfected samples contained no DNA. Crude nuclear extracts were filtered and concentrated using Microcon Ultracel YM-30 filters (Millipore, Bedford, MA) and applied directly to the PBM (Fig. 1B), except for purified samples that were immunoprecipitated from the crude extracts with the  $\alpha$ 445 antibody<sup>2</sup> (Fig. 2A) and then peptide-eluted.

### Protein Binding Microarrays

Custom 8  $\times$  15k arrays of single-stranded 42-mer to 51-mer oligonucleotides (Agilent Technologies, Santa Clara, CA) were extended on the slide in the presence of Cy3 deoxyuridine triphosphate (dUTP) using a universal primer (Fig. 1C–E) as described in Bulyk.<sup>18</sup> Both PBM1 and PBM2 contained 3000 unique sequences replicated five times, including random controls, sequences collected from the literature, mined from CHIP-chip datasets,<sup>11,19</sup> and derived from variations on the consensus 5'-AGGTCAaAGGTCA-3'. PBM2 contained sequences derived from PBM1 and sequences predicted by SVM1 on human promoter regions and the regions reported in CHIP-chip<sup>11</sup> (for a complete list of sequences on PBM1 and PBM2, see Supporting Tables 2A and 2B, respectively). Briefly, PBMs were premoistened, incubated with HNF4 $\alpha$  protein for 1 hour, washed, and then incubated with the indicated antibodies. All washes and incubations were performed at room temperature (27°C). PBMs were scanned using a GenePix Axon 4000B scanner (Molecular Devices, Sunnyvale, CA) at 543 nm (Cy3) dUTP and 633 nm (Cy5-conjugated secondary antibody). Signals were gradient-corrected using Micro-Array NORMALization of array-Comparative Genomic Hybridization data (MANOR) implemented in R.<sup>20</sup> Cross-array and intra-array normalization was performed using quantile normalization,<sup>21</sup> enabling comparison between independent experiments. Replicates for each probe were averaged, and only probes with a coefficient of variation less than 0.3 were used to train the SVM.

### SVM Training and Binding Sequence Analysis

The Kernel-based SVM (KSVM) function from Kernlab package in R with Laplace dot kernel was used to train the model (SVM1) in the classification mode<sup>22</sup> using results averaged from independent PBM1 experiments. SVM1 was then used to generate sequences for PBM2. Another SVM model in the regression mode was trained on the results of the PBM2 experiments (SVM2). For a complete list of sequences in the SVM1 and SVM2 training data, see Supporting Tables 4A and 4B, respectively. The human genome (University of California Santa Cruz [UCSC] Human Genome Browser, UCSC hg18) was searched with the binding sequences from PBM2 and the predicted binding sequences from SVM2 using the sliding window approach.

## RNA Interference and Expression Profiling Analysis

RNA interference (RNAi) against HNF4 $\alpha$ 2 was performed in HepG2 cells using small, interfering RNAs (siRNAs) corresponding to nucleotides +179 to +197 of human *HNF4A* (NM\_178849, sense siRNA: 5'-UGUGCAGGUGUUGACGAUGdTdT-3', antisense siRNA 5'-CAUCGUCAACACCUGCACAdTdT-3') (Dharmacon, Lafayette, CO). Total RNA was extracted with Trizol (Life Technologies, Carlsbad, CA) and reverse transcribed with the Reverse Transcription System (Promega, Madison, WI). Polymerase chain reaction (PCR) amplification was performed in the linear range (see Supporting Table 3B for a list of PCR primers). Expression profiling analysis was performed with Affymetrix oligonucleotide arrays (HG\_U133 Plus 2.0) using RNA from control (PGL3 siRNA) or treated (HNF4 $\alpha$  siRNA) HepG2 cells, and analyzed as previously described.<sup>13</sup>

## Chromatin Immunoprecipitation and ChIP-Chip Analysis

ChIP for HNF4 $\alpha$  from HepG2 cells on the Ninjurin 1 (*NINJI*) promoter was carried out as previously described.<sup>23</sup> HNF4 $\alpha$  ChIP-chip data from primary human hepatocytes<sup>11</sup> were extracted from ArrayExpress database, reanalyzed with the Bioconductor package LIMMA and ACME,<sup>24,25</sup> and subsequently visualized using Integrated Genome Browser (IGB; Affymetrix, Santa Clara, CA).

## Results

### Protein-Binding Microarrays Using Full-Length HNF4 $\alpha$ in Crude Nuclear Extracts

PBMs are a high-throughput *in vitro* DNA binding assay that allow for the examination of TF binding to thousands of unique sequences in a single experiment.<sup>26</sup> Recently, PBMs have been used to define the DNA-binding specificity of large classes of TFs<sup>27,28</sup> and have been shown to correlate well with gel shift results.<sup>29</sup> Whereas as others have pioneered the technology using the DNA-binding domain (DBD) of TFs purified from bacteria, here we adapt the PBM technology to more closely approximate physiological conditions. Because HNF4 $\alpha$  has a very strong dimerization domain outside of the DBD and a very low affinity for DNA when expressed in bacteria,<sup>14,30,31</sup> we ectopically expressed full-length, native HNF4 $\alpha$  in COS-7 cells and prepared minimally processed nuclear extracts (Fig. 1B) that we then applied directly to a PBM specifically designed for HNF4 $\alpha$  (Fig. 1C,D). The PBM was developed with a highly specific antibody to the C-terminus of HNF4 $\alpha$  (Supporting Fig. 1), allowing us to examine a completely native TF. The full-length HNF4 $\alpha$  protein in the crude extracts yielded an excellent signal with a range of intensities, whereas extracts from mock-transfected cells yielded no reproducible signals (Fig. 1E).

### Reproducibility and Utility of Adapted Protein-Binding Microarrays

We compared two species (rat and human) and two isoforms of HNF4 $\alpha$  (HNF4 $\alpha$ 2 and HNF4 $\alpha$ 8), as well as antibodies that recognized different regions of HNF4 $\alpha$  (Fig. 2A). There was an excellent correlation between replicate arrays in the first-generation PBM (PBM1) using crude nuclear extracts, regardless of antibody used ( $R^2 = 0.78$ ), and results with affinity-purified protein were very similar to those with crude extracts ( $R^2 = 0.68$ ) (Fig. 2B). In a second generation of the PBM (PBM2), different HNF4 $\alpha$  isoforms (HNF4 $\alpha$ 2 versus HNF4 $\alpha$ 8) and species (human versus rat) also produced excellent correlations ( $R^2 > 0.9$ ), indicating that these isoform and species differences do not influence the binding of HNF4 $\alpha$  to DNA. This is not surprising considering that the DBD is identical in these constructs (Fig. 2A).

## Accuracy of PBM and SVM

PBM1 identified ~500 new HNF4 $\alpha$  binding sequences with the DR1-derived sequences exhibiting the best binding affinities relative to negative controls ( $P < 8.274 \times 10^{-12}$ ) (Fig. 3A). Sequences derived from ChIP-chip analysis bound roughly as well as the DR1 variants. In PBM2, an additional ~1000 novel sequences that strongly bind HNF4 $\alpha$  were identified, including sequences identified by SVM1. The signal-to-noise ratio (literature-derived versus random sites) was also significantly improved in PBM2 due to optimization of the binding conditions ( $P < 2.6 \times 10^{-11}$  versus  $P < 2.6 \times 10^{-16}$ , respectively, using the Student  $t$  test) (Fig. 3B). The PBM2 results also correlated very well with gel shift results (Fig. 3C). Additionally, SVM2 derived from PBM2 predicted binding sequences with a high degree of accuracy ( $R^2 = 0.76$ ) (Fig. 3D)

## Identification of New “Rules” for HNF4 $\alpha$ DNA Binding by PBM

Even though position weight matrices (PWMs) do not capture the interdependence between the positions in a motif as do PBMs and SVMs, they are useful for describing motifs. Interestingly, the PWM of the ~450 sequences that yielded the greatest binding intensity in PBM2 (“strong binders”) did not strictly follow the DR1 rule of AGGTCA $\times$ AGGTCA. Rather, a core sequence of CAAAG is the most prominent feature, with the classical AGGTCA half-site evident only on the 3' side (Fig. 4A), a finding supported by the recent crystallographic structure of the HNF4 $\alpha$  DBD on DNA in which fewer hydrogen bonds were observed between the HNF4 $\alpha$  protein and the 5' half site.<sup>32</sup> In the PWMs for the medium and weak binding motifs, the three A's in the core appeared less frequently.

Using ~1400 strong HNF4 $\alpha$ -binding sequences obtained from PBM2, we determined the distribution of potential HNF4 $\alpha$ -binding sites in the human genome and found a broad distribution of sites with an enrichment within ~1 kilobase (kb) of the transcription start site (+1) (Fig. 4B). This is in contrast to profiles of sites for some other TFs, such as Sp1 and ELK1, that are found more exclusively near +1,<sup>33</sup> but is consistent with the fact that there are many well-characterized HNF4 $\alpha$  sites far from +1. We also found a small percentage (<1%) of sites that bound HNF4 $\alpha$  well in PBM2 but did not contain the CAAAG core (see Supporting Fig. 7 for the PWM and gel shift assay), but the biological relevance of these sequences remains to be verified.

## Expression Profiling of an HNF4 $\alpha$ RNAi Knockdown in Hepatic Cells

To identify functional HNF4 $\alpha$  target genes, we used RNAi to knock down HNF4 $\alpha$  expression in HepG2 cells, a human hepatocellular carcinoma cell line that expresses endogenous HNF4 $\alpha$  and many liver-specific genes (Fig. 5A, top panels and Supporting Fig. 5). Using the SVM2 model, we predicted several other potential HNF4 $\alpha$  target genes and determined that they were also down-regulated by reverse transcription PCR (*APOC4*, *RDH16*, *APOM*, *APOH*, *SPSB2*, *UBD*, *ZDHHC11*) (Fig. 5A, bottom panel). Whole-genome expression profiling identified ~1500 additional genes that were down-regulated (see Supporting Table 3A for a complete list). Interestingly, the gene that was down-regulated the most—Ninjurin 1 (*NINJI*) (12.5-fold)—is not a gene typically associated with HNF4 $\alpha$  function (i.e., intermediary metabolism); rather, it is involved in regulating the cell cycle. In order to determine whether *NINJI* is a direct target of HNF4 $\alpha$ , we used SVM2 to identify a potential HNF4 $\alpha$  binding site within the *NINJI* promoter region (Fig. 5B) and subsequently verified that it was bound by HNF4 $\alpha$  *in vivo* using a ChIP assay (Fig. 5C) and *in vitro* using a gel shift assay (Fig. 5D); these results suggest that *NINJI* is indeed a direct target of HNF4 $\alpha$ .



## Gene Ontology Analysis Reveals Complementary Nature of PBM, Expression Profiling, and ChIP Analysis

To compare the different methods of predicting target genes, we performed Gene Ontology (GO) on the HNF4 $\alpha$  targets predicted by RNAi expression profiling and the PBM2 search (-2 kb to +1 kb), as well as on published HNF4 $\alpha$  ChIP-chip results from primary human hepatocytes<sup>11</sup> (Fig. 6). In general, six broad biological processes contained significant GO terms for all three assays—metabolism, transport, development, regulation of signal transduction, protein modification, and apoptosis—showing the overlapping nature of the three assays. There were three additional categories—inflammatory response, cell cycle, and nucleic acid metabolism—in which genes from at least one but not all three assays were over-represented. The most notable difference between the PBM2 search from the other assays was an enrichment of genes involved in developmental processes. This is consistent with the known role of HNF4 $\alpha$  in early development,<sup>34</sup> and could be explained by the fact that the cells used in the ChIP-chip and RNAi assays are from adult stages, not embryonic stages. In general, the ChIP assay yielded more significant GO terms in all categories, which is most likely a reflection of the more specific nature of this assay and the stringent cutoff values used.

## Identification of New HNF4 $\alpha$ Target Genes and New Functions

In order to more closely compare the three methods of identifying potential target genes, we cross-referenced the PBM2 search results with the HNF4 $\alpha$  RNAi and ChIP-chip results. We identified 198 genes that were positive in all three categories, i.e., bound by HNF4 $\alpha$  in ChIP-chip, down-regulated by HNF4 $\alpha$  in HepG2 RNAi, and containing one or more verified HNF4 $\alpha$ -binding sites in the -2 kb to +1 kb region of the promoter (Fig. 7A). A similar analysis with the SVM2 search yielded 135 genes (Fig. 7B). Among these two categories, there were ~260 nonredundant genes, of which ~240 were not in the original list of HNF4 $\alpha$  target genes from the literature (Supporting Table 1A). Several of these genes are new targets within known categories of HNF4 $\alpha$  targets (e.g., homeostasis = solute carrier proteins, *SLC* genes; lipid metabolism = e.g., *ABCC6*, *DGAT2*, hydroxysteroid dehydrogenase [HSDs] genes), or more recently identified targets of HNF4 $\alpha$  (e.g., *CREB3L3*, *NR1I2*, *NR1H4*, *DOI*).<sup>35–38</sup> There were also many genes that, like *NINJI*, are in completely new categories of genes not typically associated with HNF4 $\alpha$  (e.g., signal transduction, immune response, stress response, apoptosis, cancer related, and cell structure) (Fig. 7C), several of which are reminiscent of the new functional categories identified by GO (Fig. 6). In order to determine whether the ChIP signal overlapped with the PBM or SVM sites in these new targets, all three datasets were visualized using Integrated Genome Browser. Although not all ChIP signals aligned exactly with the PBM or SVM sites, a very large number did; a sampling of these are shown in Fig. 8.

## Discussion

Identification of TF binding sites and target genes can be a laborious process. Recent genome-scale technologies such as expression profiling and genome-wide location analysis can greatly expand the repertoire of potential targets with relative ease, although the question remains as to which are direct targets that contain bona fide binding sites. PBMs allow for a high-throughput identification of DNA binding sequences that can then be integrated with the other techniques, and can also be used to predict potential new targets in additional tissues or developmental stages.

Here, we successfully adapt the PBM technology to assess HNF4 $\alpha$  DNA binding under conditions that more closely approximate physiological conditions (i.e., native full-length receptor in a crude nuclear extract) (Fig. 1). We show that the PBM results are highly

reproducible across different species (human and rat) and isoforms ( $\alpha 2$  and  $\alpha 8$ ) of HNF4 $\alpha$  under a variety of conditions (Figs. 2 and 3). We identify new rules for DNA binding and develop an SVM model to predict additional sites (Figs. 3 and 4). We compare the PBM and SVM results to RNAi expression profiling (Fig. 5) as well as to published ChIP-chip results in order to develop an integrated approach for the identification of human HNF4 $\alpha$  target genes. We show that all three systems yield similar overrepresented categories of target genes (Fig. 6), supporting the notion that specific TF binding sites in promoter regions are a major factor in driving gene expression. Using this integrated approach, we identified ~240 new, direct targets of HNF4 $\alpha$ , many of which are in new functional categories (Figs. 7 and 8). To our knowledge, this is the first such integration of extensive PBM, ChIP-chip, and expression profiling data for any TF. Finally, to facilitate future HNF4 $\alpha$  target gene research, we have developed a publicly available web-based tool (HNF4 Motif Finder) based on our PBM results that can be used to search any DNA sequence for potential HNF4 $\alpha$ -binding sites (<http://nrmotif.ucr.edu>).

We define direct targets as genes that meet three criteria: contain a functional binding site in a regulatory region (PBM/SVM search), bind *in vivo* to the promoter (ChIP), and are down-regulated when HNF4 $\alpha$  expression is knocked down (RNAi). Applying these criteria, we expand upon the classical roles of HNF4 $\alpha$  by identifying additional target genes involved in metabolism (e.g., *APOM*, *LIPC*, *LPINI*), solute carrier transport (e.g., *SLC7A2*, *SLC12A7*, *SLC25A20*), protein transport and secretion (e.g., *COPA*, *GOLGB1*, *GOLGA1*), as well as transcription regulation (e.g., *HDAC6*, *MED14*, etc.).

The integrated approach also identified new HNF4 $\alpha$  targets in pathways not previously associated with HNF4 $\alpha$ , such as regulation of signal transduction (e.g., *TAOK3*, *NGEF*, *PRKCZ*, *FNTB*), and inflammation and immune response (e.g., *IL32*, *BRE*, *LEAP2*, *IFITM2*, *BAT3*). Perhaps the most intriguing new categories of HNF4 $\alpha$  target genes are those involved in apoptosis, DNA repair, and cancer. HNF4 $\alpha$  has long been considered a key factor in hepatocyte differentiation<sup>3,4</sup> but there are an increasing number of reports indicating that HNF4 $\alpha$  may act as a tumor suppressor.<sup>39,40</sup> This view is supported by the new target genes identified here, such as *NINJI* (Fig. 5), which may play a role in regulating cellular senescence by inducing the expression of p21, a cell cycle inhibitor gene,<sup>41</sup> and is consistent with our previous findings that the p21 gene (*CDKN1A*) itself is a direct target of HNF4 $\alpha$ .<sup>23</sup> Other new HNF4 $\alpha$  target genes related to anti-growth effects are: *CIDEC*, which induces fragmentation of DNA upon apoptosis; *ATPIF1*, which inhibits an adenosine triphosphatase involved in angiogenesis; and *STEAP3*, which is induced by tumor suppressor p53 and whose down-regulation is associated with a transition from cirrhosis to hepatocellular carcinoma.<sup>42</sup> There were also genes involved in stress responses such as the DNA repair gene *FANCF*, a Fanconi's anemia complementation group F, and *USPI*, a ubiquitin-specific protease.

In addition to the genes that meet the three criteria mentioned above, our analysis also revealed thousands of additional genes that met only one or two of the three criteria. While technical considerations (e.g., missing tiles in the ChIP-chip, malfunctioning probes in the expression arrays, false positives in the ChIP assay, etc.) are sure to account for some of those genes, other explanations are also possible. For example, the genes present only in the expression profiling could be indirect targets of HNF4 $\alpha$  and hence yield no PBM/SVM or ChIP signal. Genes present in ChIP-chip alone could contain as-yet unidentified HNF4 $\alpha$ -binding sites or recruit HNF4 $\alpha$  in a non-direct fashion; it should also be noted that in Fig. 7B, we imposed a fairly stringent requirement of four or more SVM sites for a gene to be included in that analysis. Genes identified only in the PBM/SVM searches could contain *bona fide* HNF4 $\alpha$ -binding sites but are simply not expressed in the hepatocellular carcinoma cell line (HepG2) used in the expression profiling nor in the particular set of primary human

hepatocytes used in the ChIP-chip. It could also be that in adult hepatocytes the promoter regions of those genes are not available for binding (and hence activation) due to the structure of the chromatin. Genes found only in the PBM/SVM searches could also represent nonhepatic targets that are expressed in other HNF4 $\alpha$ -expressing tissues such as kidney, pancreas, intestine, and colon. Finally, it is also possible that there may be potential HNF4 $\alpha$ -binding sites in the human genome that are never used by HNF4 $\alpha$ .

Whatever the reasons for the incomplete overlap between the three assays, the use of the PBM/SVM results presented here, as well as the web-based HNF4 Motif Finder, should greatly facilitate any future investigation of potential HNF4 $\alpha$  target genes. Additionally, our approach of integrating data from multiple genome-wide assays, including PBMs, provides a powerful new framework for identifying direct targets of TFs.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

This work was funded by grants to F.M.S. (National Institutes of Health [NIH] DK053892), T.J. (National Science Foundation IIS-0711129), F.M.S. and T.J. (University of California Riverside Institute for Integrative Genome Biology, NIH R21MH087397), E.B. (PhRMA Foundation predoctoral fellowship), and W.H.-V. (University of California Toxic Substance Training Grant). We would also like to thank the following for help: A. Karatzoglou (ksvm), S. Davis (ACME), and J. Schnabl (Supporting Table 1A).

## Abbreviations

<b>ChIP</b>	chromatin immunoprecipitation
<b>DBD</b>	DNA-binding domain
<b>GO</b>	gene ontology
<b>HNF4<math>\alpha</math></b>	hepatocyte nuclear factor 4 alpha
<b>PBM</b>	protein-binding microarray
<b>PCR</b>	polymerase chain reaction
<b>PWM</b>	position weight matrix
<b>RNAi</b>	RNA interference
<b>siRNA</b>	small interfering RNA
<b>SVM</b>	Support Vector Machine
<b>TF</b>	transcription factor

## References

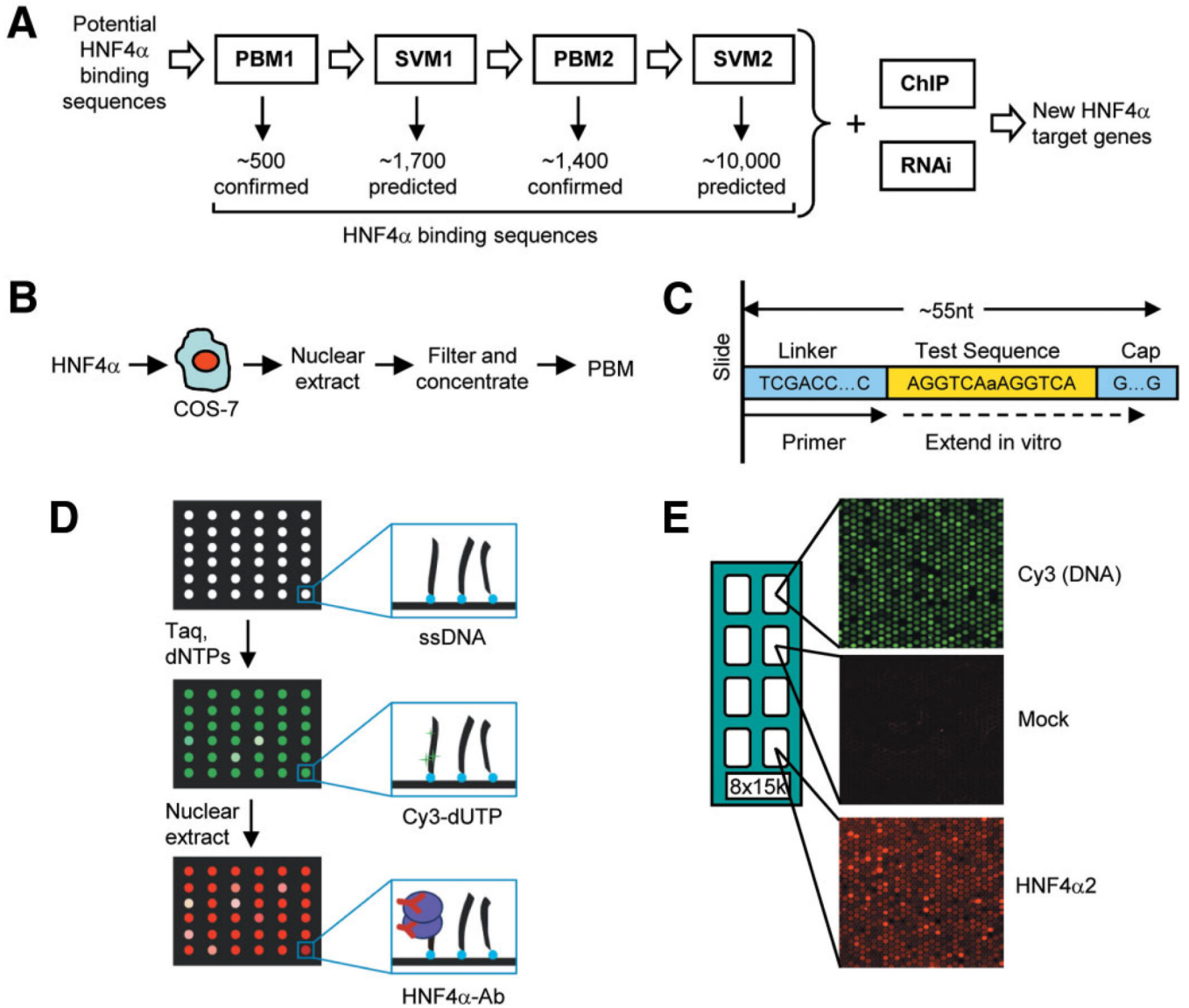
1. Bolotin E, Schnabl J, Sladek F. HNF4A (Homo sapiens). Transcription Factor Encyclopedia. 2009 <http://www.cisreg.ca/tfe>.
2. Sladek FM, Zhong WM, Lai E, Darnell JE Jr. Liver-enriched transcription factor HNF-4 is a novel member of the steroid hormone receptor superfamily. *Genes Dev.* 1990; 4:2353–2365. [PubMed: 2279702]
3. Hayhurst GP, Lee YH, Lambert G, Ward JM, Gonzalez FJ. Hepatocyte nuclear factor 4alpha (nuclear receptor 2A1) is essential for maintenance of hepatic gene expression and lipid homeostasis. *Mol Cell Biol.* 2001; 21:1393–1403. [PubMed: 11158324]



4. Watt AJ, Garrison WD, Duncan SA. HNF4: a central regulator of hepatocyte differentiation and function. *Hepatology*. 2003; 37:1249–1253. [PubMed: 12774000]
5. Gupta RK, Kaestner KH. HNF-4alpha: from MODY to late-onset type 2 diabetes. *Trends Mol Med*. 2004; 10:521–524. [PubMed: 15519277]
6. Reijnen MJ, Sladek FM, Bertina RM, Reitsma PH. Disruption of a binding site for hepatocyte nuclear factor 4 results in hemophilia B Leyden. *Proc Natl Acad Sci U S A*. 1992; 89:6300–6303. [PubMed: 1631121]
7. Sladek, F.; Seidel, S. Hepatocyte nuclear factor 4 $\alpha$ . In: Burriss, T.; McCabe, E., editors. *Nuclear Receptors and Genetic Diseases*. London: Academic Press; 2001. p. 309-361.
8. Ellard S, Colclough K. Mutations in the genes encoding the transcription factors hepatocyte nuclear factor 1 alpha (HNF1A) and 4 alpha (HNF4A) in maturity-onset diabetes of the young. *Hum Mutat*. 2006; 27:854–869. [PubMed: 16917892]
9. Gonzalez FJ. Regulation of hepatocyte nuclear factor 4 alpha-mediated transcription. *Drug Metab Pharmacokinet*. 2008; 23:2–7. [PubMed: 18305369]
10. Odom DT, Zizlsperger N, Gordon DB, Bell GW, Rinaldi NJ, Murray HL, et al. Control of pancreas and liver gene expression by HNF transcription factors. *Science*. 2004; 303:1378–1381. [PubMed: 14988562]
11. Odom DT, Dowell RD, Jacobsen ES, Nekludova L, Rolfe PA, Danford TW, et al. Core transcriptional regulatory circuitry in human hepatocytes. *Mol Syst Biol*. 2006; 2 2006.0017.
12. Rada-Iglesias A, Wallerman O, Koch C, Ameur A, Enroth S, Clelland G, 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. [PubMed: 16221759]
13. Yuan X, Ta TC, Lin M, Evans JR, Dong Y, Bolotin E, et al. Identification of an endogenous ligand bound to a native orphan nuclear receptor. *PLoS ONE*. 2009; 4:e5609. [PubMed: 19440305]
14. Bogan AA, Dallas-Yang Q, Ruse MD Jr, Maeda Y, Jiang G, Nepomuceno L, et al. Analysis of protein dimerization and ligand binding of orphan receptor HNF4alpha. *J Mol Biol*. 2000; 302:831–851. [PubMed: 10993727]
15. 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. [PubMed: 7651430]
16. Jiang G, Sladek FM. The DNA binding domain of hepatocyte nuclear factor 4 mediates cooperative binding to DNA and heterodimerization with the retinoid X receptor alpha. *J Biol Chem*. 1997; 272:1218–1225. [PubMed: 8995424]
17. Ellrott K, Yang C, Sladek FM, Jiang T. Identifying transcription factor binding sites through Markov chain optimization. *Bioinformatics*. 2002; 18(Suppl. 2):S100–S109. [PubMed: 12385991]
18. Bulyk ML. Analysis of sequence specificities of DNA-binding proteins with protein binding microarrays. *Methods Enzymol*. 2006; 410:279–299. [PubMed: 16938556]
19. Rada-Iglesias A, Wallerman O, Koch C, Ameur A, Enroth S, Clelland G, 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. [PubMed: 16221759]
20. Neuvial P, Hupe P, Brito I, Liva S, Manie E, Brennetot C, et al. Spatial normalization of array-CGH data. *BMC Bioinformatics*. 2006; 7:264. [PubMed: 16716215]
21. 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. [PubMed: 12538238]
22. Karatzoglou A, Smola A, Hornik K, Zeileis A. Kernlab—An S4 Package for Kernel Methods in R. *J Stat Softw*. 2004; 11(9) [www.jstatsoft.org](http://www.jstatsoft.org).
23. Hwang-Verslues WW, Sladek FM. Nuclear receptor hepatocyte nuclear factor 4alpha1 competes with oncoprotein c-Myc for control of the p21/WAF1 promoter. *Mol Endocrinol*. 2008; 22:78–90. [PubMed: 17885207]
24. Smyth GK, Speed T. Normalization of cDNA microarray data. *Methods*. 2003; 31:265–273. [PubMed: 14597310]

25. Scacheri PC, Crawford GE, Davis S. Statistics for ChIP-chip and DNase hypersensitivity experiments on NimbleGen arrays. *Methods Enzymol.* 2006; 411:270–282. [PubMed: 16939795]
26. Berger MF, Bulyk ML. Universal protein-binding microarrays for the comprehensive characterization of the DNA-binding specificities of transcription factors. *Nat Protoc.* 2009; 4:393–411. [PubMed: 19265799]
27. Berger MF, Badis G, Gehrke AR, Talukder S, Philippakis AA, Pena-Castillo L, et al. Variation in homeodomain DNA binding revealed by high-resolution analysis of sequence preferences. *Cell.* 2008; 133:1266–1276. [PubMed: 18585359]
28. Badis G, Berger MF, Philippakis AA, Talukder S, Gehrke AR, Jaeger SA, et al. Diversity and complexity in DNA recognition by transcription factors. *Science.* 2009; 324:1720–1723. [PubMed: 19443739]
29. Linnell J, Mott R, Field S, Kwiatkowski DP, Ragoussis J, Udalova IA. Quantitative high-throughput analysis of transcription factor binding specificities. *Nucleic Acids Res.* 2004; 32:e44. [PubMed: 14990752]
30. Jiang G, Lee U, Sladek FM. Proposed mechanism for the stabilization of nuclear receptor DNA binding via protein dimerization. *Mol Cell Biol.* 1997; 17:6546–6554. [PubMed: 9343418]
31. Jiang G, Nepomuceno L, Yang Q, Sladek FM. Serine/threonine phosphorylation of orphan receptor hepatocyte nuclear factor 4. *Arch Biochem Biophys.* 1997; 340:1–9. [PubMed: 9126270]
32. Lu P, Rha GB, Melikishvili M, Wu G, Adkins BC, Fried MG, et al. Structural basis of natural promoter recognition by a unique nuclear receptor, HNF4alpha. *Diabetes gene product.* *J Biol Chem.* 2008; 283:33685–33697. [PubMed: 18829458]
33. Yang C, Bolotin E, Jiang T, Sladek FM, Martinez E. Prevalence of the initiator over the TATA box in human and yeast genes and identification of DNA motifs enriched in human TATA-less core promoters. *Gene.* 2007; 389:52–65. [PubMed: 17123746]
34. Lemaigre F, Zaret KS. Liver development update: new embryo models, cell lineage control, and morphogenesis. *Curr Opin Genet Dev.* 2004; 14:582–590. [PubMed: 15380251]
35. Luebke-Wheeler J, Zhang K, Battle M, Si-Tayeb K, Garrison W, Chhinder S, et al. Hepatocyte nuclear factor 4alpha is implicated in endoplasmic reticulum stress-induced acute phase response by regulating expression of cyclic adenosine monophosphate responsive element binding protein H. *Hepatology.* 2008; 48:1242–1250. [PubMed: 18704925]
36. Kamiya A, Inoue Y, Gonzalez FJ. Role of the hepatocyte nuclear factor 4alpha in control of the pregnane X receptor during fetal liver development. *Hepatology.* 2003; 37:1375–1384. [PubMed: 12774017]
37. Ohguchi H, Tanaka T, Uchida A, Magoori K, Kudo H, Kim I, et al. Hepatocyte nuclear factor 4alpha contributes to thyroid hormone homeostasis by cooperatively regulating the type 1 iodothyronine deiodinase gene with GATA4 and Kruppel-like transcription factor 9. *Mol Cell Biol.* 2008; 28:3917–3931. [PubMed: 18426912]
38. Zhang Y, Castellani LW, Sinal CJ, Gonzalez FJ, Edwards PA. Peroxisome proliferator-activated receptor-gamma coactivator 1alpha (PGC-1alpha) regulates triglyceride metabolism by activation of the nuclear receptor FXR. *Genes Dev.* 2004; 18:157–169. [PubMed: 14729567]
39. Tanaka T, Jiang S, Hotta H, Takano K, Iwanari H, Sumi K, et al. Dys-regulated expression of P1 and P2 promoter-driven hepatocyte nuclear factor-4alpha in the pathogenesis of human cancer. *J Pathol.* 2006; 208:662–672. [PubMed: 16400631]
40. Yin C, Lin Y, Zhang X, Chen YX, Zeng X, Yue HY, et al. Differentiation therapy of hepatocellular carcinoma in mice with recombinant adenovirus carrying hepatocyte nuclear factor-4alpha gene. *Hepatology.* 2008; 48:1528–1539. [PubMed: 18925631]
41. Toyama T, Sasaki Y, Horimoto M, Iyoda K, Yakushijin T, Ohkawa K, et al. Ninjurin1 increases p21 expression and induces cellular senescence in human hepatoma cells. *J Hepatol.* 2004; 41:637–643. [PubMed: 15464245]
42. Caillot F, Daveau R, Daveau M, Lubrano J, Saint-Auret G, Hiron M, et al. Down-regulated expression of the TSAP6 protein in liver is associated with a transition from cirrhosis to hepatocellular carcinoma. *Histopathology.* 2009; 54:319–327. [PubMed: 19236508]
43. Crooks GE, Hon G, Chandonia JM, Brenner SE. WebLogo: a sequence logo generator. *Genome Res.* 2004; 14:1188–1190. [PubMed: 15173120]

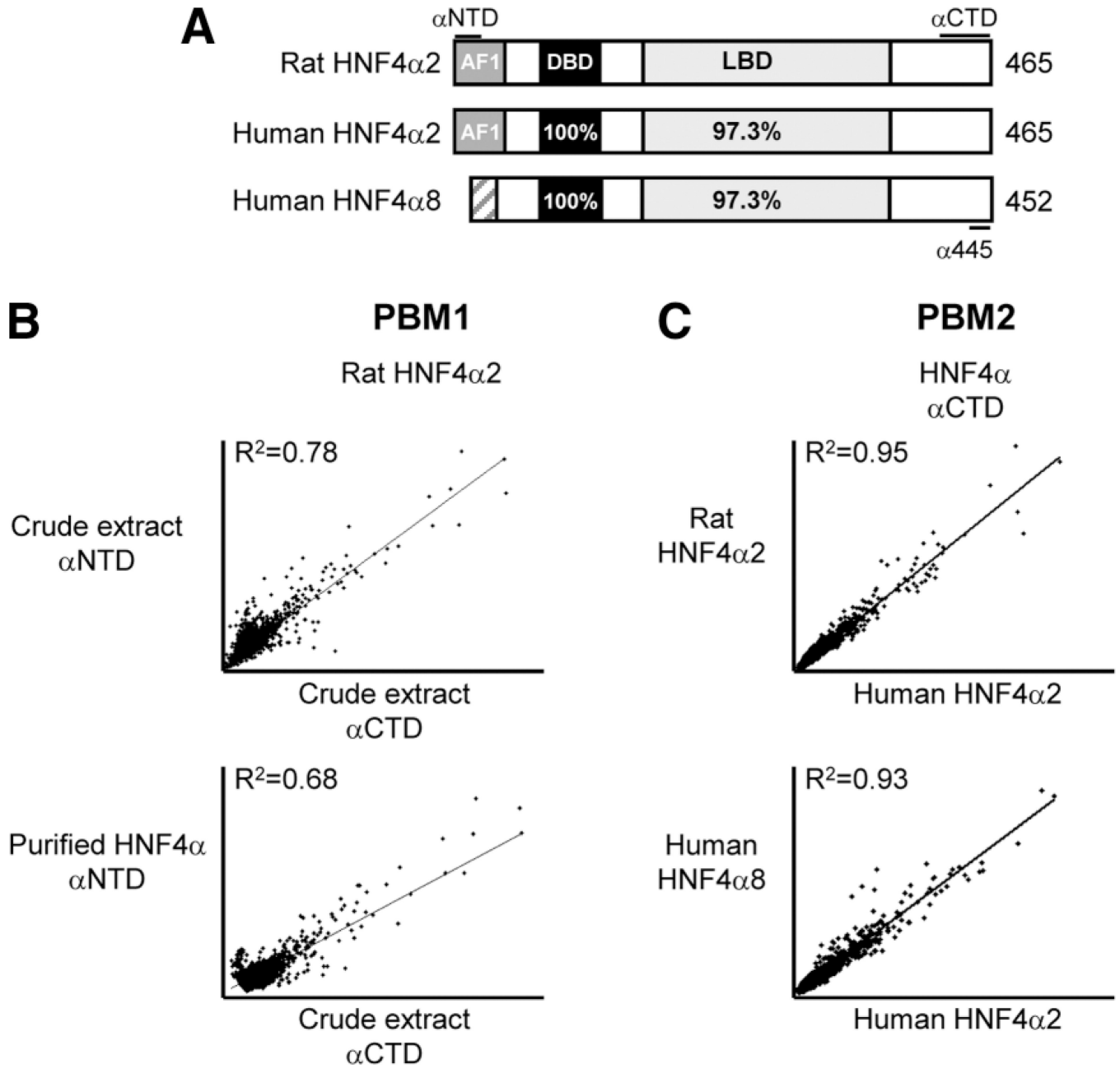
44. Dennis G Jr, Sherman BT, Hosack DA, Yang J, Gao W, Lane HC, et al. DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biol.* 2003; 4:P3. [PubMed: 12734009]



**Fig. 1.** Integrated approach for the identification of direct target genes and protein binding microarray (PBM) design. (A) Overview of workflow. Known and predicted HNF4 $\alpha$ -binding sequences (217 sequences from the literature, sites predicted by the Markov model and ChIP-chip analysis, and random controls) were printed on the first-generation PBM (PBM1) and incubated with minimally processed crude nuclear extracts from COS-7 cells transfected with full-length HNF4 $\alpha$  (B). Results from the initial screen were used to train the Support Vector Machine (SVM1), resulting in 1700 predicted HNF4 $\alpha$ -binding sequences that were printed onto a second-generation PBM (PBM2). Searches of human promoters using PBM/SVM results were cross-referenced with results from RNAi expression profiling and ChIP-chip to identify new HNF4 $\alpha$  targets. (C, D) Overview of PBM. Single-stranded oligonucleotides with a common linker, test sequence, and a G/C-rich cap region (C) printed on the PBM were extended *in vitro* in the presence of Cy3-dUTP (D). The PBMs were incubated with extracts containing HNF4 $\alpha$  and visualized by immunofluorescence. (E) Typical PBM results are shown. Double-stranded DNA with Cy3 incorporated (top panel), mock-transfected cells lacking HNF4 $\alpha$  (middle panel), and

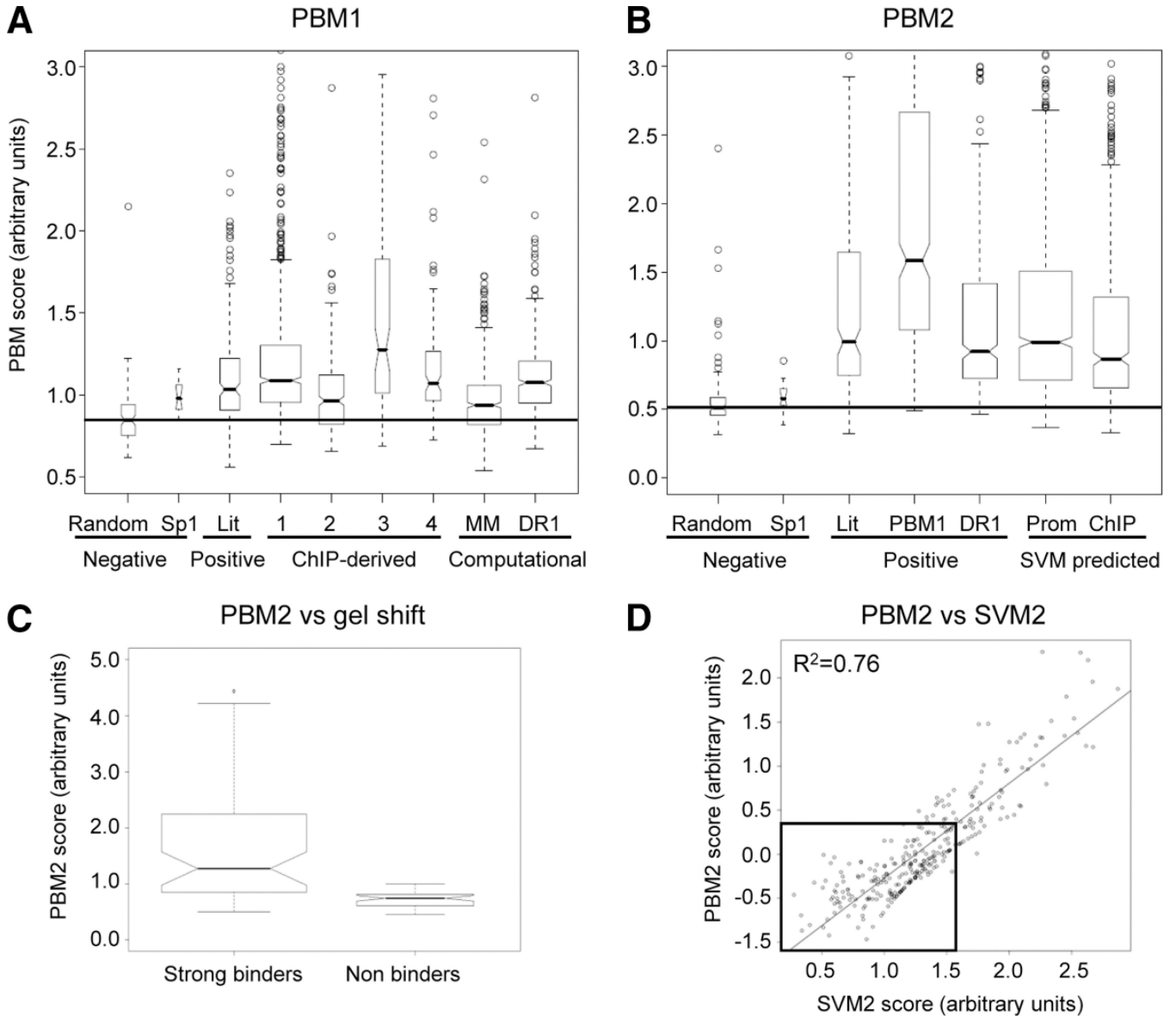
extracts containing HNF4 $\alpha$  with fluorescent signal proportional to the binding affinity (bottom panel). An array of 8  $\times$  15k, Agilent microarray slide with eight replicate subarrays with ~3000 unique sequences each spotted five times (~15,000 spots) per subarray. Supporting Fig. 1 shows that nontransfected COS-7 cells do not express HNF4 $\alpha$  and that the antibody used to detect HNF4 $\alpha$  in the PBM is completely specific. Supporting Fig. 2 shows a linear relationship between Cy3 incorporation and the number of A's in the extended sequence.





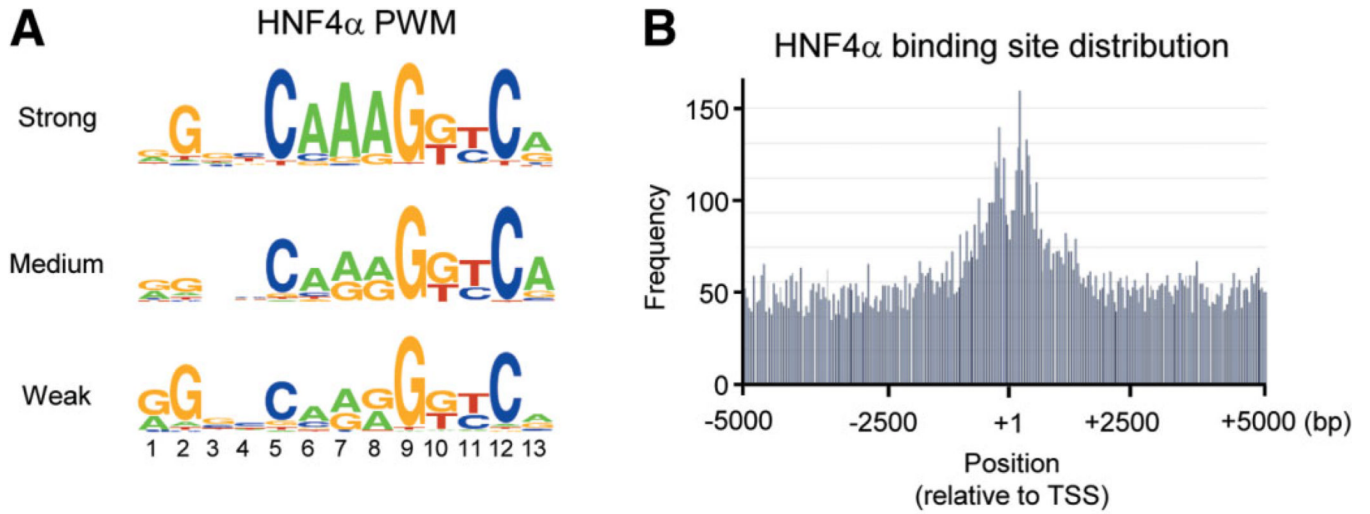
**Fig. 2.** Reproducibility of the PBM. (A) Diagram of HNF4 $\alpha$  splice variants used in PBM indicating percent amino acid identity in conserved regions. AF1, activation function 1; DBD, DNA-binding domain; LBD, ligand-binding domain. The regions of the protein detected by the monoclonal antibodies ( $\alpha$ NTD, amino-terminal HNF4 $\alpha$  antibody;  $\alpha$ CTD, carboxy-terminal HNF4 $\alpha$  antibody) and the affinity-purified polyclonal antibody  $\alpha$ 445 are indicated. (See Supporting Materials and Methods for additional details on plasmids and antibodies.) (B) Scatter plot of individual spot intensities showing correlation between PBM1 using rat HNF4 $\alpha$ 2 protein and the  $\alpha$ NTD and  $\alpha$ CTD antibodies (top panel) as well as purified HNF4 $\alpha$ 2 versus crude nuclear extracts (bottom panel). (C) Scatter plot of PBM2 results as

in (B) comparing different HNF4 $\alpha$  isoforms from different species. See Supporting Figs. 3 and 4 for scatter plot matrices of PBM1 and PBM2 from nine experiments.



**Fig. 3.** Relative binding affinities of HNF4 $\alpha$ -binding sites. (A) Box plot of sequence categories represented on PBM1 and corresponding PBM score averaged from six independent arrays with each sequence spotted five times. Box width indicates the relative number of sequences per category. Nonoverlapping box plot notches strongly indicate that the medians significantly differ ( $P < 0.05$ ). Boxes and whiskers (dashed line) represent quartiles of binding scores for each sequence category. Line, median of random sequences. Negative controls = randomly generated 13-mers; known Sp1 sites derived from the literature. Positive controls = 217 known HNF4 $\alpha$ -binding sites from the literature (Lit) (Supporting Tables 1A and 1B). ChIP-derived, binding sites derived from published HNF4 $\alpha$  ChIP-chip data: 1, from Odom et al.<sup>11</sup>; 2, from Rada-Iglesias et al.<sup>12</sup>; 3, our analysis of Odom et al. data using Bioproprospector software; 4, our analysis of Odom et al. data using AlignACE software. Computational, binding sequences derived from our permuted Markov model (MM)<sup>17</sup> and permutations of the DR1 consensus sequence (DR1). (B) Box plot of sequence categories represented in PBM2 (three independent arrays) as in (A). PBM1, best 500

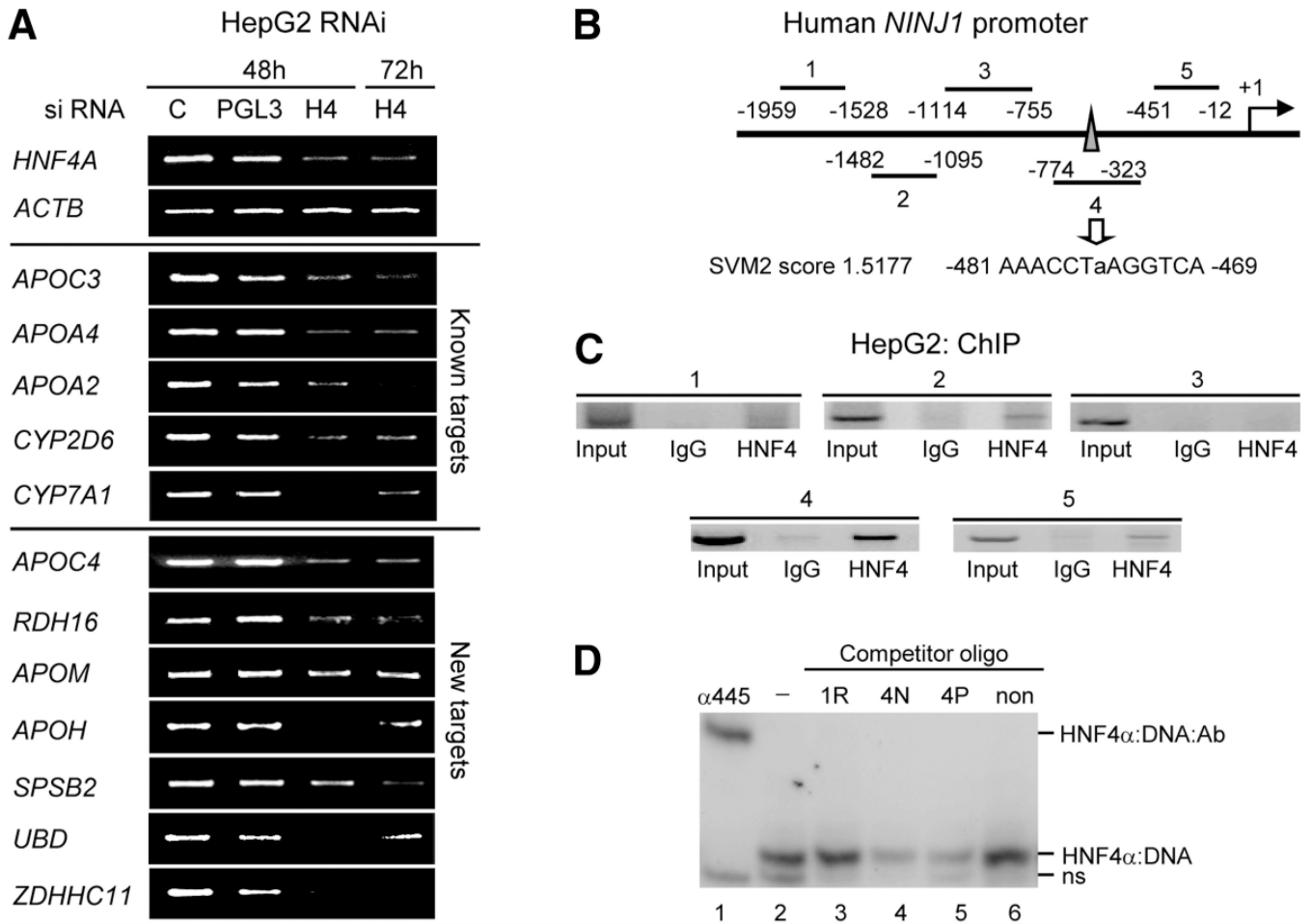
sequences from PBM1; SVM predicted, sequences from SVM1 search of promoter regions of all annotated human genes (Prom) and ChIP-chip data (ChIP).<sup>11</sup> For a complete list of all the sequences on PBM1 and PBM2 and binding scores, see Supporting Tables 2A and 2B. (C) Box plot of PBM2 results versus results from ~100 gel shift experiments showing a statistically significant difference (Student *t* test,  $P < 0.00622$ ) between strong binders and nonbinders or very weak binders (see Supporting Materials and Methods and Supporting Fig. 6 for results). (D) Scatter plot of log(PBM2) intensity compared to SVM2 score of one of the 10-fold cross validation results used to evaluate the predictive power of SVM2. A cutoff of an SVM2 score  $>1.51$ , corresponding to three standard deviations from the mean of random controls, was used to identify binding sequences in subsequent analyses.



**Fig. 4.**

Position weight matrix (PWM) for HNF4 $\alpha$ -binding sequence motif and HNF4 $\alpha$ -binding site distribution. (A) PWM of HNF4 $\alpha$ -binding sequences derived from PBM2. All sequences with relative binding affinity at least 2 standard deviations above the mean of the random controls were divided into three groups of ~450 each—strong, medium and weak—and used to generate the PWMs.<sup>43</sup> (B) Distribution of potential HNF4 $\alpha$ -binding sites around the transcription start site (TSS, +1) of all human promoters (UCSC hg18) as determined by an exact match search with PBM2 results. Sites are overrepresented in the -1 kb to +1 kb region. (See Supporting Fig. 7 for PWM and gel shifts of noncanonical binding sites detected in the PBM.)





**Fig. 5.** HNF4 $\alpha$  knockdown in HepG2 cells using RNAi and identification of Ninjurin 1 as a direct target of HNF4 $\alpha$ . (A) Verification of HNF4 $\alpha$ 1/2 knockdown. HepG2 cells treated with siRNA for the hours indicated. Reverse transcription PCR was performed on the indicated HNF4 $\alpha$  targets. C, no siRNA. PGL3, control siRNA. H4, HNF4 $\alpha$  siRNA (all splice variants from the P1 promoter are targeted). (B) Human *NINJ1* promoter showing regions amplified by PCR in ChIP in (C). Region 4 contains a predicted HNF4 $\alpha$ -binding site with an SVM2 score of ~1.5177 (moderate binding affinity). (C) ChIP result of HNF4 $\alpha$  in HepG2 cells on the human *NINJ1* promoter using PCR primers that amplify regions 1–4 noted in (B). IgG, normal rabbit immunoglobulin G; HNF4,  $\alpha$ 445 antibody. (D) Gel shift assay using nuclear extracts from COS-7 cells transfected with rat HNF4 $\alpha$ 2, radiolabeled probe from the ApoA1 promoter and unlabeled competitors in 250-fold molar excess corresponding to the SVM site identified in region 4 with native flanking sequences (4N) or PBM flanking sequences (4P) as well as a known nonbinder (non, 175 TTR) and a randomly chosen sequence from region 1 (1R). Shown are the HNF4 $\alpha$ :DNA shift complex, a supershift complex with the  $\alpha$ 445 antibody (HNF4 $\alpha$ :DNA:antibody) and nonspecific band from the COS-7 extracts (ns); free probe is not shown. See Supporting Materials and Methods for details on gel shift conditions, Supporting Fig. 5 for immunoblot of HNF4 $\alpha$  protein in the RNAi, Supporting Table 3A for a complete list of genes that are down-regulated, Supporting Table 3B for primer sequences, and Supporting Table 8 for gel shift sequences.

**A** Classical Functions

Metabolism	ChIP	RNAi	PBM
metabolic process	***	***	**
alcohol biosynthetic process	***		
alcohol metabolic process	***	***	***
amine biosynthetic process	**	***	
amine catabolic process	***	*	*
amine metabolic process	***	***	**
amino acid and derivative metabolic process	***	***	
amino acid biosynthetic process	***	***	
amino acid catabolic process	***	**	*
amino acid metabolic process	***	***	*
aminoglycan metabolic process		***	*
aromatic compound metabolic process	***	**	
carbohydrate biosynthetic process	***	**	*
carbohydrate metabolic process	***	***	**
carboxylic acid metabolic process	***	***	**
catabolic process	***	***	*
cellular lipid metabolic process	***	***	***
cellular protein metabolic process	***	**	
cofactor metabolic process	***	*	
electron transport	***		
fatty acid metabolic process	***	*	*
fatty acid oxidation	***		
generation of precursor metabolites and energy	***		
glucose metabolic process	***		
glutamine family amino acid metabolic process		***	
hexose metabolic process	***	**	*
lipid biosynthetic process	***	***	**
lipid metabolic process	***	***	***
macromolecule biosynthetic process	***		
monocarboxylic acid metabolic process	***	***	*
monosaccharide biosynthetic process	***		
monosaccharide metabolic process	***	***	*
nitrogen compound metabolic process	***	***	**
organic acid metabolic process	***	***	**
pyruvate metabolic process	***		
steroid biosynthetic process	***	*	*

Transport	ChIP	RNAi	PBM
transport	***	***	***
Golgi vesicle transport	***	***	*
vesicle-mediated transport	***	***	**
di-, tri-valent inorganic cation transport			***

Classical Functions (cont.)

Development	ChIP	RNAi	PBM
developmental process			***
anatomical structure development			***
cell differentiation			***
cellular component organization and biogenesis	***	*	***
multicellular organismal development			***
organelle organization and biogenesis	***		***

**B** New Functions

Regulation of Signal Transduction	ChIP	RNAi	PBM
biological regulation	**		***
regulation of signal transduction		*	***
response to endogenous stimulus	***		*
response to organic substance	***		

Protein Modification	ChIP	RNAi	PBM
ubiquitin cycle	***	***	**
protein modification process	***	***	*

Apoptosis	ChIP	RNAi	PBM
apoptosis	***		*
negative regulation of apoptosis	***	*	**
programmed cell death	***		

Inflammatory Response	ChIP	RNAi	PBM
response to stress	***		*
acute inflammatory response	***		

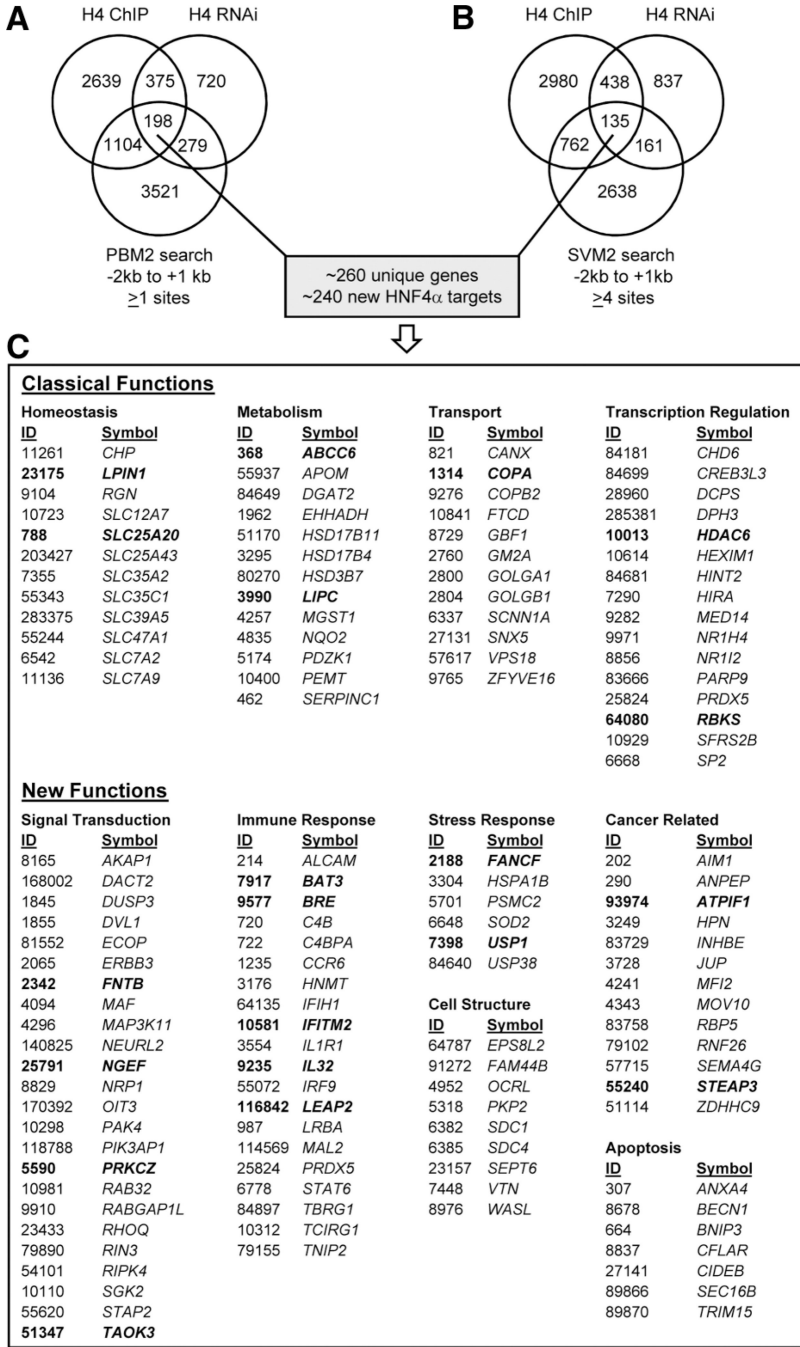
Cell Cycle	ChIP	RNAi	PBM
cell cycle	***		
regulation of cell cycle	***		*

Nucleic Acid Modification	ChIP	RNAi	PBM
DNA repair	***		
RNA processing	***		
mRNA metabolic process	***		

**Fig. 6.** Comparative Gene Ontology for genes bound *in vivo* by HNF4 $\alpha$  (ChIP-chip), down-regulated in HNF4 $\alpha$  RNAi, and containing PBM or SVM HNF4 $\alpha$  binding sites. Overrepresented categories from Gene Ontology analysis using DAVID<sup>44</sup> of HNF4 $\alpha$  ChIP-chip from primary human hepatocytes<sup>11</sup> (ChIP), expression profiling of HNF4 $\alpha$  knocked down in HepG2 cells using RNAi (RNAi) and PBM2 search of -2 kb to +1 kb of all annotated human genes (UCSC hg18) (PBM). Shown are the biological processes for which at least one of the three methods had a *P* value (EASE-score) of < 0.001 (\*\*\*), < 0.01 (\*\*), or < 0.05 (\*). Redundant categories were removed. (A) Biological processes related to classical HNF4 $\alpha$  target genes well-established in the literature (e.g., Supporting Table 1A).

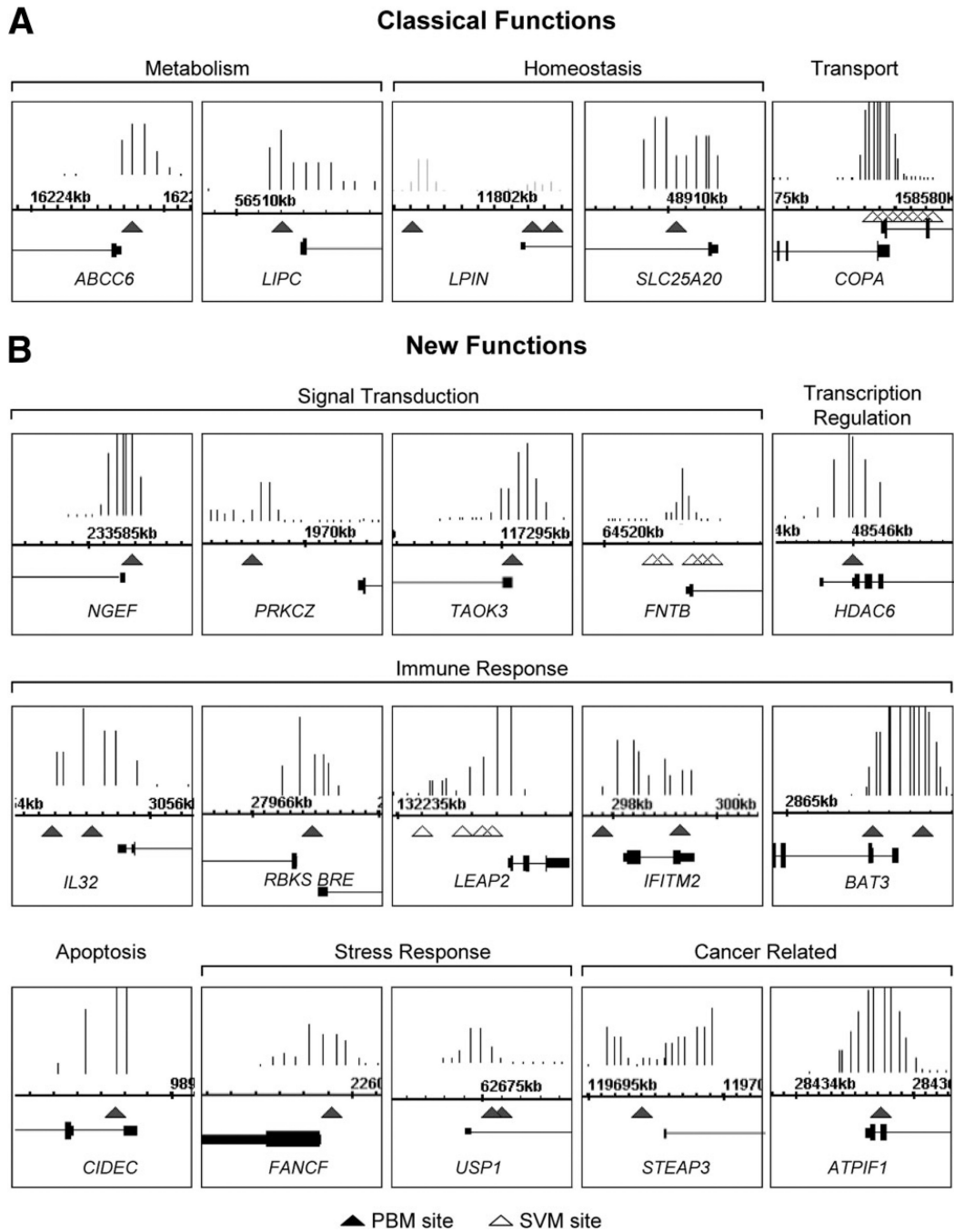
(B) Biological processes not typically associated with HNF4 $\alpha$ . See Supporting Table 5 for a complete list of GO terms and *P* values for the ChIP, PBM, and RNAi as well as the SVM search ( 4 sites in -2 kb to +1 kb).



**Fig. 7.** Cross-reference of three methods used to identify potential human HNF4α target genes: ChIP-chip, RNAi expression profiling, and PBM/SVM binding site search. (A) Venn analysis of genes: bound by HNF4α in primary human hepatocytes (H4 ChIP)<sup>11</sup>; down-regulated in expression profiling by HNF4α siRNA in HepG2 cells (H4 RNAi) (Fig. 5); and containing a potential HNF4α-binding site as determined by an exact match search using PBM2 results of annotated human genes (UCSC hg18) -2 kb to +1 kb relative to the TSS (PBM2 search). Shown are the number of genes; genes in the intersection are likely to be direct targets of HNF4α. (B) As in (A) except with SVM2 search of annotated human genes with four or more sites. (See Supporting Tables 6A and 6B for a complete list of the 198 and

135 genes in the intersection of the Venn diagrams in (A) and (B), respectively.) (C) Sampling of new HNF4 $\alpha$  target genes that are bound *in vivo*, down-regulated in HNF4 $\alpha$  knockdown, and containing 1 PBM or 4 SVM sites. Functions classically associated with HNF4 $\alpha$  are shown as well as new functional categories. ID, Entrez Gene ID; Symbol, Official Gene Symbol. (See Supporting Tables 7A and 7B for a complete listing of all human genes with one or more PBM sites and four or more SVM sites, respectively.)





**Fig. 8.** Illustration of select new HNF4 $\alpha$  target genes down-regulated in RNAi, bound *in vivo*, and with PBM or SVM HNF4 $\alpha$ -binding sites. Screenshots from Integrated Genome Browser of HNF4 $\alpha$  ChIP-chip signals from primary human hepatocytes in promoter regions<sup>11</sup> with PBM (closed triangle) sites indicated. SVM sites (open triangle) are indicated only for those genes lacking a PBM site in the region shown. ChIP signals are all statistically significant. Numbers are chromosome coordinates from UCSC hg18. Not all shots are on the same scale. Classical (A) and new functions (B) as defined in Fig. 7c are indicated.