SUPPLEMENTAL MATERIAL

"Core transcriptional regulatory circuitry of human hepatocytes"

Odom D.T. *et al. Mol. Systems Biol. (2006) submission*

List of Supplemental Figures

Figure S1. Liver master regulators bind to the promoters of the HNF4α and TTR genes
Figure S2. Binding sequences in Odom et al. 2006 and Rada-Iglesias et al. 2005.
Figure S3. Combinatorial occupancy of promoter regions is significantly enriched
Figure S4. Correspondence of gene expression and location data

List of Supplemental Tables

Table S1. Hepatocyte master regulators profiled
Table S2. Regulatory network motifs in human liver
Table S3. *S. cerevisiae* transcription factors with auto-regulatory loops
Table S4. Auto-regulation occurrence among yeast transcriptional regulators sorted into groups of regulators, based on number of genes bound
Table S5. Mammalian transcription factors reported to directly auto-regulate
Table S6. Mammalian transcription factors that do not directly auto-regulate
Table S7. Tissues, antisera, and array platforms used for ChIP experiments in Rada-Iglesias et al. 2005 and Odom et al. 2006
Table S8. Binding event overlap between Rada-Iglesias et al. 2005 and Odom et al. 2006.
Table S9. Statistical significance of sequence motifs for the six liver master regulators.

List of Supplemental Sections

- ***Auto-regulation among Saccharyomyces cerevisiae master regulators***
- ***Comparison between Rada-Iglesias ENCODE data and Odom promoter data***
- ***Materials and methods***

Website

Further experimental details are available on the authors' website at
http://web.wi.mit.edu/young/autoregulation/

Data Accession Number

Complete, unprocessed data from these experiments have been deposited into the public database ArrayExpress (http://www.ebi.ac.uk/arrayexpress/): accession: **E-WMIT-9**

Figure S1A. The HNF4α gene has an enhancer located 6 kb upstream from the P1 transcriptional start site (Bailly et al. 2001). HNF1α and HNF4α ChIP enrichment is shown as a trace relative to genomic position (graphic below). The known HNF binding sites in the HNF4A P1 promoter are located within 1 kb of the TSS and in the enhancer region 6 kb from the transcriptional start site, and are shown as color shaded ovals along the genome track. Coverage of the promoter regions was typically one 60-mer for each 250 bases of sequence, and allowed identification of in vivo binding sites to within tens of bases of the binding site position.

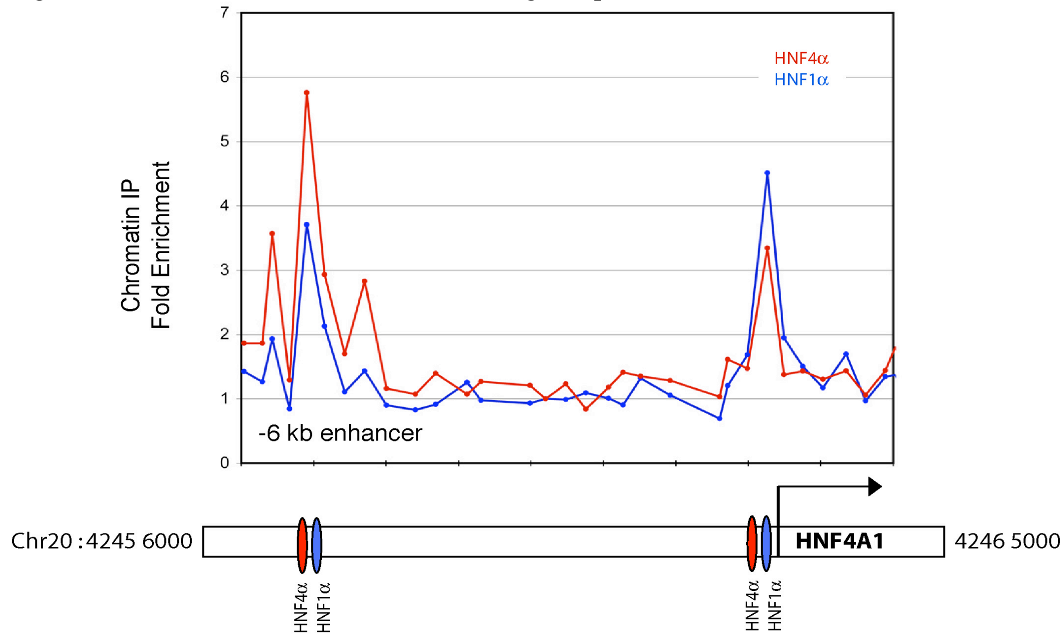


Figure S1B. The TTR gene is bound by a number of HNF factors in vivo at previously known sites within its immediate proximal promoter (reviewed in Costa et al. 2003). The known binding sites in the TTR promoter are located within 250 bp of the TSS, and are shown as color shaded ovals.

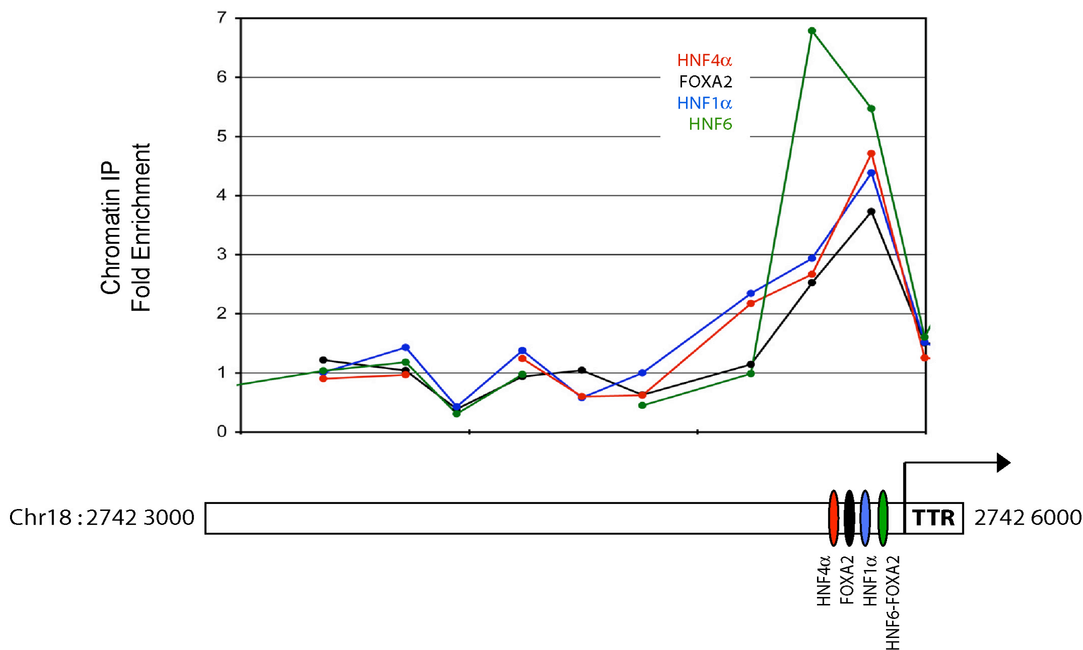


Table S1. Hepatocyte master regulators profiled in this study.*

Name	Function	Antisera	Gene Symbol	PFAM category	Accession
HNF1α	Metabolic control	sc-6547 (Santa Cruz)	TCF1	POU-homeodomain	NM_000545
HNF4α	Development, metabolism	sc-8987 (Santa Cruz)	NR2A1	Nuclear receptor	NM_008261
HNF6	Development	sc-13050 (Santa Cruz)	ONECUT1	CUT-homeodomain	NM_004498
HNF3β	Development	R. Costa	FOXA2	Forkhead	NM_021784
CREB1	Nutrient response	pCREB-5322 (Montminy)	CREB1	bZIP	NM_004379
USF1	Glucose, Lipid metabolism	AB/HLH230 (Cemines)	USF1	bHLH	NM_007122

* **References**

Cereghini, S. (1996). Liver-enriched transcription factors and hepatocyte differentiation. Faseb J 10, 267-282.

Hayashi, Y., Wang, W., Ninomiya, T., Nagano, H., Ohta, K., and Itoh, H. (1999). Liver enriched transcription factors and differentiation of hepatocellular carcinoma. Mol Pathol 52, 19-24.

Zaret, K. S. (2002). Regulatory phases of early liver development: paradigms of organogenesis. Nat Rev Genet 3, 499-512.

Costa, R. H., Kalinichenko, V. V., Holterman, A. X., and Wang, X. (2003). Transcription factors in liver development, differentiation, and regeneration. Hepatology *38*, 1331-1347.

Duncan, S. A., Navas, M. A., Dufort, D., Rossant, J., and Stoffel, M. (1998). Regulation of a transcription factor network required for differentiation and metabolism. Science 281, 692-695.

Krivan, W., and Wasserman, W. W. (2001). A predictive model for regulatory sequences directing liver-specific transcription. Genome Res 11, 1559-1566.

HNF1α

Boj, S. F., Parrizas, M., Maestro, M. A., and Ferrer, J. (2001). A transcription factor regulatory circuit in differentiated pancreatic cells. Proc Natl Acad Sci U S A 98, 14481-14486.

Hiraiwa, H., Pan, C. J., Lin, B., Akiyama, T. E., Gonzalez, F. J., and Chou, J. Y. (2001). A molecular link between the common phenotypes of type 1 glycogen storage disease and HNF1alpha-null mice. J Biol Chem 276, 7963-7967.

Pontoglio, M., Sreenan, S., Roe, M., Pugh, W., Ostrega, D., Doyen, A., Pick, A. J., Baldwin, A., Velho, G., Froguel, P., et al. (1998). Defective insulin secretion in hepatocyte nuclear factor 1alpha-deficient mice. J Clin Invest 101, 2215-2222.

Shih, D. Q., Bussen, M., Sehayek, E., Ananthanarayanan, M., Shneider, B. L., Suchy, F. J., Shefer, S., Bollileni, J. S., Gonzalez, F. J., Breslow, J. L., and Stoffel, M. (2001). Hepatocyte nuclear factor-1alpha is an essential regulator of bile acid and plasma cholesterol metabolism. Nat Genet 27, 375-382.

HNF4α

Bailly, A., Torres-Padilla, M. E., Tinel, A. P., and Weiss, M. C. (2001). An enhancer element 6 kb upstream of the mouse HNF4alpha1 promoter is activated by glucocorticoids and liver-enriched transcription factors. Nucleic Acids Res *29*, 3495-3505.

Hayhurst, G. P., Lee, Y. H., Lambert, G., Ward, J. M., and Gonzalez, F. J. (2001). Hepatocyte nuclear factor 4alpha (nuclear receptor 2A1) is essential for maintenance of hepatic gene expression and lipid homeostasis. Mol Cell Biol 21, 1393-1403.

Gupta, R. K., and Kaestner, K. H. (2004). HNF-4alpha: from MODY to late-onset type 2 diabetes. Trends Mol Med 10, 521-524.

Gupta, R. K., Vatamaniuk, M. Z., Lee, C. S., Flaschen, R. C., Fulmer, J. T., Matschinsky, F. M., Duncan, S. A., and Kaestner, K. H. (2005). The MODY1 gene HNF-4alpha regulates selected genes involved in insulin secretion. J Clin Invest 115, 1006-1015.

Lee, Y. H., Sauer, B., and Gonzalez, F. J. (1998). Laron dwarfism and non-insulin-dependent diabetes mellitus in the Hnf-1alpha knockout mouse. Mol Cell Biol 18, 3059-3068.

Parviz, F., Matullo, C., Garrison, W. D., Savatski, L., Adamson, J. W., Ning, G., Kaestner, K. H., Rossi, J. M., Zaret, K. S., and Duncan, S. A. (2003). Hepatocyte nuclear factor 4alpha controls the development of a hepatic epithelium and liver morphogenesis. Nat Genet 34, 292-296.

HNF6

Jacquemin, P., Lemaigre, F. P., and Rousseau, G. G. (2003). The Onecut transcription factor HNF-6 (OC-1) is required for timely specification of the pancreas and acts upstream of Pdx-1 in the specification cascade. Dev Biol 258, 105-116.

Lemaigre, F. P., Durviaux, S. M., Truong, O., Lannoy, V. J., Hsuan, J. J., and Rousseau, G. G. (1996). Hepatocyte nuclear factor 6, a transcription factor that contains a novel type of homeodomain and a single cut domain. Proc Natl Acad Sci U S A 93, 9460-9464.

Rausa, F., Samadani, U., Ye, H., Lim, L., Fletcher, C. F., Jenkins, N. A., Copeland, N. G., and Costa, R. H. (1997). The cut-homeodomain transcriptional activator HNF-6 is coexpressed with its target gene HNF-3 beta in the developing murine liver and pancreas. Dev Biol 192, 228-246.

Samadani, U., and Costa, R. H. (1996). The transcriptional activator hepatocyte nuclear factor 6 regulates liver gene expression. Mol Cell Biol 16, 6273-6284.

FOXA2/HNF3β

Rausa, F. M., Tan, Y., Zhou, H., Yoo, K. W., Stolz, D. B., Watkins, S. C., Franks, R. R., Unterman, T. G., and Costa, R. H. (2000). Elevated levels of hepatocyte nuclear factor 3beta in mouse hepatocytes influence expression of genes involved in bile acid and glucose homeostasis. Mol Cell Biol 20, 8264-8282.

Wolfrum, C., Asilmaz, E., Luca, E., Friedman, J. M., and Stoffel, M. (2004). Foxa2 regulates lipid metabolism and ketogenesis in the liver during fasting and in diabetes. Nature 432, 1027-1032.

Zhang, L., Rubins, N. E., Ahima, R. S., Greenbaum, L. E., and Kaestner, K. H. (2005). Foxa2 integrates the transcriptional response of the hepatocyte to fasting. Cell Metab 2, 141-148.

CREB1

Montminy, M., Koo, S. H., and Zhang, X. (2004). The CREB family: key regulators of hepatic metabolism. Ann Endocrinol (Paris) 65, 73-75.

Sharma, C., Pradeep, A., Pestell, R. G., and Rana, B. (2004). Peroxisome proliferator-activated receptor gamma activation modulates cyclin D1 transcription via beta-catenin-independent and cAMP-response element-binding protein-dependent pathways in mouse hepatocytes. J Biol Chem 279, 16927-16938.

Koo, S.H., Flechner, L., Qi, L., Zhang, X., Screaton, R.A., Jeffries, S., Hedrick, S., Xu, W., Boussouar, F., Brindle, P., Takemori, H. and Montminy, M. (2005) The CREB coactivator TORC2 is a key regulator of fasting glucose metabolism. Nature, 437, 1109-1111.

USF1

Casado, M., Vallet, V. S., Kahn, A., and Vaulont, S. (1999). Essential role in vivo of upstream stimulatory factors for a normal dietary response of the fatty acid synthase gene in the liver. J Biol Chem 274, 2009-2013.

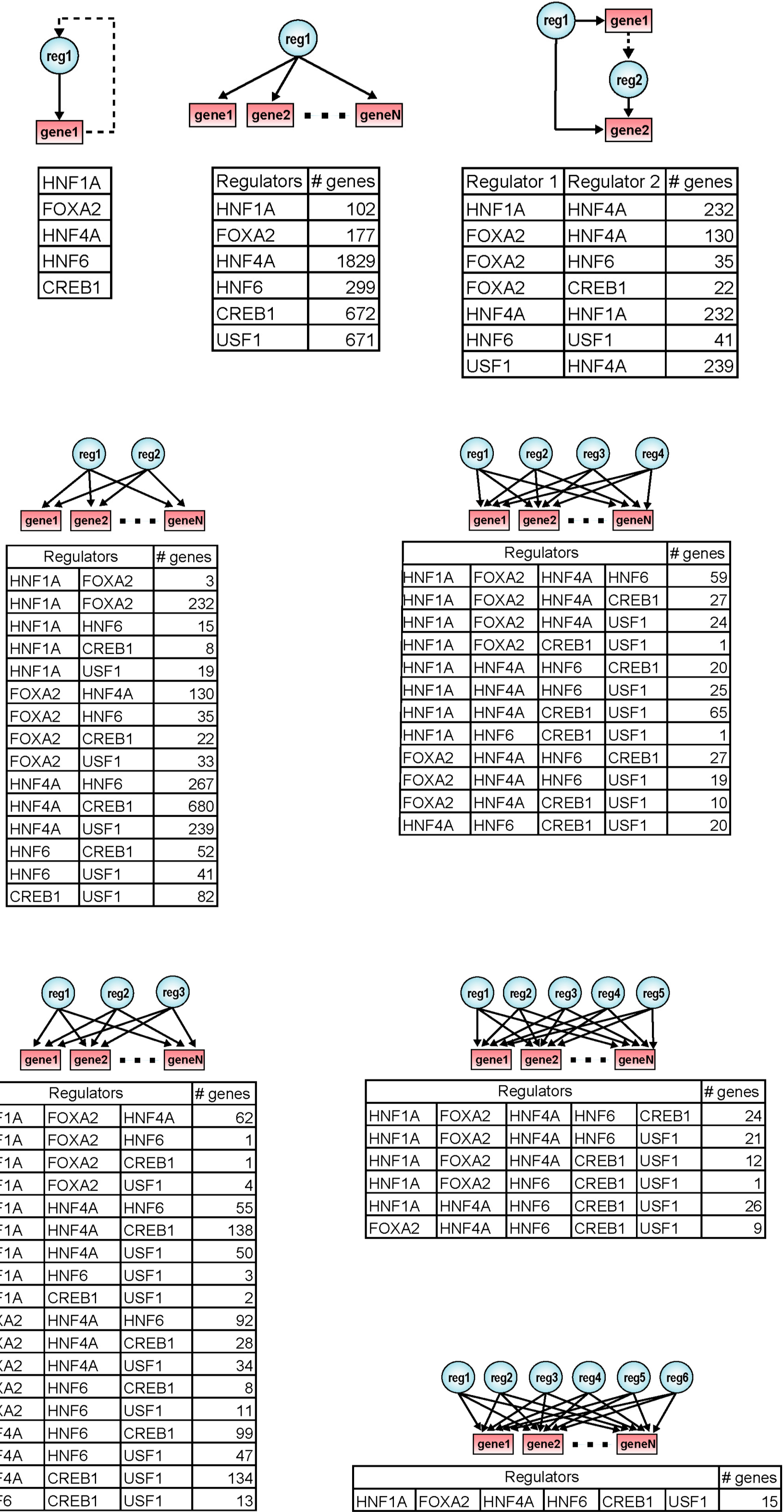
Iynedjian, P. B. (1998). Identification of upstream stimulatory factor as transcriptional activator of the liver promoter of the glucokinase gene. Biochem J 333 (Pt 3), 705-712.

Pajukanta, P., Lilja, H. E., Sinsheimer, J. S., Cantor, R. M., Lusis, A. J., Gentile, M., Duan, X. J., Soro-Paavonen, A., Naukkarinen, J., Saarela, J., et al. (2004). Familial combined hyperlipidemia is associated with upstream transcription factor 1 (USF1). Nat Genet 36, 371-376.

Ribeiro, A., Pastier, D., Kardassis, D., Chambaz, J., and Cardot, P. (1999). Cooperative binding of upstream stimulatory factor and hepatic nuclear factor 4 drives the transcription of the human apolipoprotein A-II gene. J Biol Chem 274, 1216-1225.

Vallet, V. S., Casado, M., Henrion, A. A., Bucchini, D., Raymondjean, M., Kahn, A., and Vaulont, S. (1998). Differential roles of upstream stimulatory factors 1 and 2 in the transcriptional response of liver genes to glucose. J Biol Chem 273, 20175-20179.

Table S2. Regulatory network motifs in human liver, determined as described in Materials and methods.



HNF1A
FOXA2
HNF4A
HNF6
CREB1

Regulators	# genes
HNF1A	102
FOXA2	177
HNF4A	1829
HNF6	299
CREB1	672
USF1	671

Regulator 1	Regulator 2	# genes
HNF1A	HNF4A	232
FOXA2	HNF4A	130
FOXA2	HNF6	35
FOXA2	CREB1	22
HNF4A	HNF1A	232
HNF6	USF1	41
USF1	HNF4A	239

Regulators		# genes
HNF1A	FOXA2	3
HNF1A	FOXA2	232
HNF1A	HNF6	15
HNF1A	CREB1	8
HNF1A	USF1	19
FOXA2	HNF4A	130
FOXA2	HNF6	35
FOXA2	CREB1	22
FOXA2	USF1	33
HNF4A	HNF6	267
HNF4A	CREB1	680
HNF4A	USF1	239
HNF6	CREB1	52
HNF6	USF1	41
CREB1	USF1	82

Regulators				# genes
HNF1A	FOXA2	HNF4A	HNF6	59
HNF1A	FOXA2	HNF4A	CREB1	27
HNF1A	FOXA2	HNF4A	USF1	24
HNF1A	FOXA2	CREB1	USF1	1
HNF1A	HNF4A	HNF6	CREB1	20
HNF1A	HNF4A	HNF6	USF1	25
HNF1A	HNF4A	CREB1	USF1	65
HNF1A	HNF6	CREB1	USF1	1
FOXA2	HNF4A	HNF6	CREB1	27
FOXA2	HNF4A	HNF6	USF1	19
FOXA2	HNF4A	CREB1	USF1	10
HNF4A	HNF6	CREB1	USF1	20

Regulators			# genes
HNF1A	FOXA2	HNF4A	62
HNF1A	FOXA2	HNF6	1
HNF1A	FOXA2	CREB1	1
HNF1A	FOXA2	USF1	4
HNF1A	HNF4A	HNF6	55
HNF1A	HNF4A	CREB1	138
HNF1A	HNF4A	USF1	50
HNF1A	HNF6	USF1	3
HNF1A	CREB1	USF1	2
FOXA2	HNF4A	HNF6	92
FOXA2	HNF4A	CREB1	28
FOXA2	HNF4A	USF1	34
FOXA2	HNF6	CREB1	8
FOXA2	HNF6	USF1	11
HNF4A	HNF6	CREB1	99
HNF4A	HNF6	USF1	47
HNF4A	CREB1	USF1	134
HNF6	CREB1	USF1	13

Regulators					# genes
HNF1A	FOXA2	HNF4A	HNF6	CREB1	24
HNF1A	FOXA2	HNF4A	HNF6	USF1	21
HNF1A	FOXA2	HNF4A	CREB1	USF1	12
HNF1A	FOXA2	HNF6	CREB1	USF1	1
HNF1A	HNF4A	HNF6	CREB1	USF1	26
FOXA2	HNF4A	HNF6	CREB1	USF1	9

Regulators						# genes
HNF1A	FOXA2	HNF4A	HNF6	CREB1	USF1	15

Table S3. *S. cerevisiae* transcription factors with auto-regulatory loops in rich media*

Regulator	Process
STB5	Multidrug resistance
SWI4	Cell cycle
STE12	Mating
SUM1	Sporulation
TEC1	Pseudohyphal growth
PDR3	Multidrug resistance
RCS1	Iron homeostasis
RDS1	Multidrug resistance
ROX1	An/aerobic growth
SMP1	Osmotic stress
SUT1	Hypoxia response
YAP6	Salt tolerance
YAP7	Stress response
ZAP1	Heavy metal exposure
ARG81	Arginine metablism
ARO80	Aromatic amino acid regulation
CBF1	Centromere stability
HAP1	Aerobic respiration
NRG1	Glucose regulation
RAP1	Protein biosynthesis
INO2	Phospholipid biosynthesis
IME1	Meiosis

Table S4. Auto-regulation occurrence among yeast transcriptional regulators sorted into groups of regulators based on number of genes bound.

Minimum out-degree	Cumul. Observed	Cumul. Expected	p-value	Group size
60	10	3.8	0.003	35
50	11	4.3	0.003	40
40	13	5.3	0.002	49
30	14	6.9	0.007	64
20	17	9.9	0.019	92
10	19	13.5	0.079	125
0	22	22	0.533	204

<u>Auto-regulation among S. cerevisiae master regulators</u>

A master regulator is a transcription factor that controls cellular processes using a combination of direct and indirect means. We used the yeast S. cerevisiae as a model to explore the correlation between the presence of auto-regulation and identification of a transcription factor as a master regulator. This comparison was made possible because a complete set of binding interactions is available for all known and suspected transcriptional regulators. We analyzed the two cases of direct and indirect regulation separately, though both mechanisms can be present in varying degrees. We reasoned that in order to directly control a cellular process, which can include scores of genes, master regulators as a class would be expected to bind a disproportionately large fraction of yeast promoter regions. In contrast, if a master regulator were controlling a cellular process through intermediary transcription factors, and was present at the top of a hierarchy of factors, we would expect to observe the master regulator binding to a disproportionate number of promoter regions of other transcription factors, which would then act downstream. We used these definitions to test whether auto-regulatory loops were correlated with master regulators of yeast cellular functions.

First, we explored whether auto-regulation was found disproportionately among regulators that function indirectly. To identify these master regulators based on the number of downstream transcription factors, we ranked the 204 transcriptional factors by the number of

other transcription factors whose promoters they bind in vivo, and selected the top fifth. There are 39 transcription factors in this category, and they bind to the promoters of at least 3 and at most 13 transcriptional regulators (average= 5.0). This set of 39 captured 11 auto-regulatory events, for a p-value of 4.9×10^{-4}.

Second, we analyzed whether auto-regulation is more common among master regulators that function directly, by binding many genes. Using a variety of thresholds for the number of bound genes from ten to sixty, we found that auto-regulation is more common among the master regulators than would be expected (Table S4). For this calculation, we assumed a null hypothesis that the probability of auto-regulation was the same for all regulators (P=22/204), and thus independent of the number of genes bound by the regulator (out-degree). We also tested an alternative null hypothesis in which the probability of auto-regulation was equal to the out-degree divided by the total number of bound genes. This calculation yielded similar results (not shown).

*** <u>References</u>**

Ashburner, B. P., and Lopes, J. M. (1995). Auto-regulated expression of the yeast INO2 and INO4 helix-loop-helix activator genes effects cooperative regulation on their target genes. Mol Cell Biol 15, 1709-1715.

Delahodde, A., Delaveau, T., and Jacq, C. (1995). Positive auto-regulation of the yeast transcription factor Pdr3p, which is involved in control of drug resistance. Mol Cell Biol 15, 4043-4051.

Shefer-Vaida, M., Sherman, A., Ashkenazi, T., Robzyk, K., and Kassir, Y. (1995). Positive and negative feedback loops affect the transcription of IME1, a positive regulator of meiosis in Saccharomyces cerevisiae. Dev Genet 16, 219-228.

Ren, B., Robert, F., Wyrick, J. J., Aparicio, O., Jennings, E. G., Simon, I., Zeitlinger, J., Schreiber, J., Hannett, N., Kanin, E., et al. (2000). Genome-wide location and function of DNA binding proteins. Science 290, 2306-2309.

Eiznhamer, D. A., Ashburner, B. P., Jackson, J. C., Gardenour, K. R., and Lopes, J. M. (2001). Expression of the INO2 regulatory gene of Saccharomyces cerevisiae is controlled by positive and negative promoter elements and an upstream open reading frame. Mol Microbiol 39, 1395-1405.

Simon, I., Barnett, J., Hannett, N., Harbison, C. T., Rinaldi, N. J., Volkert, T. L., Wyrick, J. J., Zeitlinger, J., Gifford, D. K., Jaakkola, T. S., and Young, R. A. (2001). Serial regulation of transcriptional regulators in the yeast cell cycle. Cell 106, 697-708.

Lee, T. I., Rinaldi, N. J., Robert, F., Odom, D. T., Bar-Joseph, Z., Gerber, G. K., Hannett, N. M., Harbison, C. T., Thompson, C. M., Simon, I., et al. (2002). Transcriptional regulatory networks in Saccharomyces cerevisiae. Science 298, 799-804.

Zeitlinger, J., Simon, I., Harbison, C. T., Hannett, N. M., Volkert, T. L., Fink, G. R., and Young, R. A. (2003). Program-specific distribution of a transcription factor dependent on partner transcription factor and MAPK signaling. Cell 113, 395-404.

Harbison, C. T., Gordon, D. B., Lee, T. I., Rinaldi, N. J., Macisaac, K. D., Danford, T. W., Hannett, N. M., Tagne, J. B., Reynolds, D. B., Yoo, J., et al. (2004). Transcriptional regulatory code of a eukaryotic genome. Nature 431, 99-104.

Table S5. Vertebrate transcriptional regulators with an auto-regulatory loop**

Factor	Symbol	First Author	Year	Species	Tissue
cREL	REL	Hannink	1990	Chicken	
MYB	MYB	Nicolaides	1991	Hampster	Fibroblasts
AR	AR	Grad	1999	Human	Prostate
BCL6	BCL6	Wang	2002	Human	B-cells
CEBPA	CEBPA	Timchenko	1995	Human	Liver
CEBPA	CEBPA	Foka	2001	Human	Hepatoma
CEBPB	CEBPB	Niehof	2001	Human	Hepatoma
c-jun	JUN	Angel	1988	Human	Cell lines
c-jun	JUN	Mechta	1994	Human	Review
C-MYC	MYC	Martinotti	1988	Human	B-cells
C-MYC	MYC	Kitaura	1991	Human	HeLa
CREB	CREB1	Meyer	1993	Human	Placental cells
E2F1	E2F1	Neuman	1994	Human	Osteogenic Sarcoma
ER	ESR1	Castles	1997	Human	HeLa
ETS1	ETS1	Jorcyk	1991	Human	HeLa
ETS1	ETS1	Oka	1991	Human	Lymphocyte
ETS1	ETS1	Majerus	1992	Human	T-cell
GATA3	GATA3	Zhou	2003	Human	Th1/Th2
GFI1	GFI1	Doan	2004	Human	Lymphocyte
junD	JUND	Berger	1994	Human	HeLa
junD	JUND	Berger	1998	Human	HeLa
LXRA	NR1H3	Lafitte	2001	Human	Lymphocyte
LXRA	NR1H3	Li	2002	Human	Liver
MITF-M	MITF	Saito	2002	Human	HeLa
NANOG	NANOG	Boyer	2005	Human	Embryonic stem cells
N-Myc	MYCN	Sivak	1997	Human	Neuroblastoma
p49/p100	NFKB2	Lombardi	1995	Human	HeLa
p50	NFKB1	Ten	1992	Human	Cell lines
p50	NFKB1	Cogswell	1993	Human	Cell lines
P53	TP53	Hudson	1995	Human	Cell lines
P53	TP53	Benoit	2000	Human	Adenocarcinoma
PAX6	PAX6	Yamaguchi	1997	Human	Ocular
PU.1	SPI1	Chen	1995	Human	Myeloid
PURA	PURA	Muralidharan	2001	Human	Glial/other
RB1	RB1	Ohtani-Fujita	1994	Human	Cell lines
RB1	RB1	Shan	1994	Human	Osteogenic Sarcoma
RBL1	RBL1	Zhu	1995	Human	Embryonic Kidney
SOX2	SOX2	Boyer	2005	Human	Embryonic stem cells
TTF1	TITF1	Oguchi	1998	Human	Thyroid
KLF4	KLF4	Mahatan	1999	Monkey	Kidney
BRN3A	POU4F1	Trieu	2003	Mouse	Neurons
CDX1	CDX1	Beland	2004	Mouse	Embryonic carcinoma
CDX2	CDX2	Xu	1999	Mouse	Pancreas
CEBPA	CEBPA	Christy	1991	Mouse	Adipocytes
CEBPD	CEBPD	Yamada	1999	Mouse	
c-Fos	FOS	Sassone-Corsi	1988	Mouse	Embryonic Fibroblasts
c-Fos	FOS	Konig	1989	Mouse	Embryonic Fibroblasts
CRY1	CRY1	Kume	1999	Mouse	Adipocytes
FOXA2	FOXA2	Pani	1992	Mouse	Liver
GSC	GSC	Danilov	1998	Mouse	Embryonic carcinoma

Table S5 (cont'd). Vertebrate transcriptional regulators with an auto-regulatory loop**

HDAC1	HDAC1	Schuettengruber	2003	Mouse	Fibroblasts
HES1	HES1	Hirata	2002	Mouse	Myoblasts/Fibroblasts/Neuroblastoma/
HES7	HES7	Bessho	2001	Mouse	
HNF4A	HNF4A	Briancon	2004	Mouse	Liver
LXRA	NR1H3	Ulven	2004	Mouse	Adipose
MYOD	MYOD	Tapscott	1991	Mouse	Muscle
MYOD	MYOD1	Blais	2005	Mouse	Myoblast
MYOD	MYOD1	Thayer	1989	Mouse	Embryonic Fibroblasts
MYOG	MYOG1	Blais	2005	Mouse	Myoblast
OCT1	POU2F1	Pankratova	2003	Mouse	Fibroblasts
OCT4	POU5F1	Okura-Nakanishi	2005	Mouse	mES
P53	TP53	Deffie	1993	Mouse	Embryonic Fibroblasts
PAX4	PAX4	Smith	2000	Mouse	Beta-cell line
PAX6	PAX6	Aota	2003	Mouse	Ocular
PIT1	POU1F1	Chen	1990	Mouse	Epithelial
PITX1	PITX1	Goodyer	2003	Mouse	Stomodeum/Oral ectoderm
PU.1	SPI1	Okuno	2005	Mouse	Bone Marrow
RB1	RB1	Hamel	1992	Mouse	Embryonic carcinoma
RUNX2	RUNX2	Drissi	2000	Mouse	Osteoblasts
RUNX2	RUNX2	Tou	2003	Mouse	Osteoblasts
SOX2	SOX2	Catena	2004	Mouse	mES
SRF	SRF	Spencer	1996	Mouse	Embryonic Fibroblasts
STAT3	STAT3	Narimatsu	2001	Mouse	T-cell
STRA13	STRA13	Sun	2000	Mouse	Embryonic Fibroblasts
TR	THRA	Sadow	2003	Mouse	Liver/Heart
EGR1	EGR1	Saadane	2000	Mouse	Cardiocytes
GATA2	GATA2	Grass	2003	Mouse	Hematopoietic
NKX2.5	NKX2-5	Oka	1997	Mouse	Cardiocytes
CEBPA	CEBPA	Legraverend	1993	Rat	Liver
CEBPD	CEBPD	Tanabe	2000	Rat	
CREB	CREB1	Walker	1995	Rat	Testicular/Sertoli
CREM	CREM	Molina	1993	Rat	Cell lines
E2F2	E2F2	Sears	1997	Rat	Embryonic Fibroblasts
FRA1	FOSL1	Casalino	2003	Rat	Thyroid
N-Myc	MYCN	Facchini	1997	Rat	Fibroblasts
PIT1	POU1F1	Theill	1993	Rat	Pituitary
xVENT2B	xVENT2B	Henningfeld	2002	Xenopus	Embryos
GATA1	GATA1	Nishikawa	2003	Zebrafish	Hematopoietic
GATA1	GATA1	Tsai	1991	Mouse	Hematopoietic
GATA1	GATA1	Orkin	1995	Mouse	Hematopoietic

** **References**

Angel, P., Hattori, K., Smeal, T. and Karin, M. (1988) The jun proto-oncogene is positively auto-regulated by its product, Jun/AP-1. Cell, 55, 875-885.

Aota, S., Nakajima, N., Sakamoto, R., Watanabe, S., Ibaraki, N. and Okazaki, K. (2003) Pax6 auto-regulation mediated by direct interaction of Pax6 protein with the head surface ectoderm-specific enhancer of the mouse Pax6 gene. Dev Biol, 257, 1-13.

Beland, M., Pilon, N., Houle, M., Oh, K., Sylvestre, J.R., Prinos, P. and Lohnes, D. (2004) Cdx1 auto-regulation is governed by a novel Cdx1-LEF1 transcription complex. Mol Cell Biol, 24, 5028-5038.

Benoit, V., Hellin, A.C., Huygen, S., Gielen, J., Bours, V. and Merville, M.P. (2000) Additive effect between NF-kappaB subunits and p53 protein for transcriptional activation of human p53 promoter. Oncogene, 19, 4787-4794.

Berger, I. and Shaul, Y. (1994) The human junD gene is positively and selectively auto-regulated. DNA Cell Biol, 13, 249-255.
Berger, I. and Shaul, Y. (1998) c-Fos antagonizes the junD gene positive auto-regulatory loop; a novel c-Fos role in promoter switching. Gene, 211, 375-382.
Bessho, Y., Sakata, R., Komatsu, S., Shiota, K., Yamada, S. and Kageyama, R. (2001) Dynamic expression and essential functions of Hes7 in somite segmentation. Genes Dev, 15, 2642-2647.
Blais, A., Tsikitis, M., Acosta-Alvear, D., Sharan, R., Kluger, Y. and Dynlacht, B.D. (2005) An initial blueprint for myogenic differentiation. Genes Dev, 19, 553-569.
Boyer, L.A., Lee, T.I., Cole, M.F., Johnstone, S.E., Levine, S.S., Zucker, J.P., Guenther, M.G., Kumar, R.M., Murray, H.L., Jenner, R.G., Gifford, D.K., Melton, D.A., Jaenisch, R. and Young, R.A. (2005) Core transcriptional regulatory circuitry in human embryonic stem cells. Cell, 122, 947-956.
Briancon, N., Bailly, A., Clotman, F., Jacquemin, P., Lemaigre, F.P. and Weiss, M.C. (2004) Expression of the alpha7 isoform of hepatocyte nuclear factor (HNF) 4 is activated by HNF6/OC-2 and HNF1 and repressed by HNF4alpha1 in the liver. J Biol Chem, 279, 33398-33408.
Casalino, L., De Cesare, D. and Verde, P. (2003) Accumulation of Fra-1 in ras-transformed cells depends on both transcriptional auto-regulation and MEK-dependent posttranslational stabilization. Mol Cell Biol, 23, 4401-4415.
Castles, C.G., Oesterreich, S., Hansen, R. and Fuqua, S.A. (1997) Auto-regulation of the estrogen receptor promoter. J Steroid Biochem Mol Biol, 62, 155-163.
Catena, R., Tiveron, C., Ronchi, A., Porta, S., Ferri, A., Tatangelo, L., Cavallaro, M., Favaro, R., Ottolenghi, S., Reinbold, R., Scholer, H. and Nicolis, S.K. (2004) Conserved POU binding DNA sites in the Sox2 upstream enhancer regulate gene expression in embryonic and neural stem cells. J Biol Chem, 279, 41846-41857.
Chen, R.P., Ingraham, H.A., Treacy, M.N., Albert, V.R., Wilson, L. and Rosenfeld, M.G. (1990) Auto-regulation of pit-1 gene expression mediated by two cis-active promoter elements. Nature, 346, 583-586.
Christy, R.J., Kaestner, K.H., Geiman, D.E. and Lane, M.D. (1991) CCAAT/enhancer binding protein gene promoter: binding of nuclear factors during differentiation of 3T3-L1 preadipocytes. Proc Natl Acad Sci U S A, 88, 2593-2597.
Cogswell, P.C., Scheinman, R.I. and Baldwin, A.S., Jr. (1993) Promoter of the human NF-kappa B p50/p105 gene. Regulation by NF-kappa B subunits and by c-REL. J Immunol, 150, 2794-2804.
Cripps, R.M., Lovato, T.L. and Olson, E.N. (2004) Positive auto-regulation of the Myocyte enhancer factor-2 myogenic control gene during somatic muscle development in Drosophila. Dev Biol, 267, 536-547.
Danilov, V., Blum, M., Schweickert, A., Campione, M. and Steinbeisser, H. (1998) Negative auto-regulation of the organizer-specific homeobox gene goosecoid. J Biol Chem, 273, 627-635.
Deffie, A., Wu, H., Reinke, V. and Lozano, G. (1993) The tumor suppressor p53 regulates its own transcription. Mol Cell Biol, 13, 3415-3423.
Doan, L.L., Porter, S.D., Duan, Z., Flubacher, M.M., Montoya, D., Tsichlis, P.N., Horwitz, M., Gilks, C.B. and Grimes, H.L. (2004) Targeted transcriptional repression of Gfi1 by GFI1 and GFI1B in lymphoid cells. Nucleic Acids Res, 32, 2508-2519.
Drissi, H., Luc, Q., Shakoori, R., Chuva De Sousa Lopes, S., Choi, J.Y., Terry, A., Hu, M., Jones, S., Neil, J.C., Lian, J.B., Stein, J.L., Van Wijnen, A.J. and Stein, G.S. (2000) Transcriptional auto-regulation of the bone related CBFA1/RUNX2 gene. J Cell Physiol, 184, 341-350.
Facchini, L.M., Chen, S., Marhin, W.W., Lear, J.N. and Penn, L.Z. (1997) The Myc negative auto-regulation mechanism requires Myc-Max association and involves the c-myc P2 minimal promoter. Mol Cell Biol, 17, 100-114.
Fitch, M.J., Donato, J.J. and Tye, B.K. (2003) Mcm7, a subunit of the presumptive MCM helicase, modulates its own expression in conjunction with Mcm1. J Biol Chem, 278, 25408-25416.
Foka, P., Kousteni, S. and Ramji, D.P. (2001) Molecular characterization of the Xenopus CCAAT-enhancer binding protein beta gene promoter. Biochem Biophys Res Commun, 285, 430-436.
Goodyer, C.G., Tremblay, J.J., Paradis, F.W., Marcil, A., Lanctot, C., Gauthier, Y. and Drouin, J. (2003) Pitx1 in vivo promoter activity and mechanisms of positive auto-regulation. Neuroendocrinology, 78, 129-137.
Grad, J.M., Dai, J.L., Wu, S. and Burnstein, K.L. (1999) Multiple androgen response elements and a Myc consensus site in the androgen receptor (AR) coding region are involved in androgen-mediated up-regulation of AR messenger RNA. Mol Endocrinol, 13, 1896-1911.

Grass, J.A., Boyer, M.E., Pal, S., Wu, J., Weiss, M.J. and Bresnick, E.H. (2003) GATA-1-dependent transcriptional repression of GATA-2 via disruption of positive auto-regulation and domain-wide chromatin remodeling. Proc Natl Acad Sci U S A, 100, 8811-8816.
Hamel, P.A., Gill, R.M., Phillips, R.A. and Gallie, B.L. (1992) Transcriptional repression of the E2-containing promoters EIIaE, c-myc, and RB1 by the product of the RB1 gene. Mol Cell Biol, 12, 3431-3438.
Hannink, M. and Temin, H.M. (1990) Structure and auto-regulation of the c-rel promoter. Oncogene, 5, 1843-1850.
Hatzis, P. and Talianidis, I. (2001) Regulatory mechanisms controlling human hepatocyte nuclear factor 4alpha gene expression. Mol Cell Biol, 21, 7320-7330.
Henningfeld, K.A., Friedle, H., Rastegar, S. and Knochel, W. (2002) Auto-regulation of Xvent-2B; direct interaction and functional cooperation of Xvent-2 and Smad1. J Biol Chem, 277, 2097-2103.
Hirata, H., Yoshiura, S., Ohtsuka, T., Bessho, Y., Harada, T., Yoshikawa, K. and Kageyama, R. (2002) Oscillatory expression of the bHLH factor Hes1 regulated by a negative feedback loop. Science, 298, 840-843.
Hudson, J.M., Frade, R. and Bar-Eli, M. (1995) Wild-type p53 regulates its own transcription in a cell-type specific manner. DNA Cell Biol, 14, 759-766.
Ingraham, H.A., Chen, R.P., Mangalam, H.J., Elsholtz, H.P., Flynn, S.E., Lin, C.R., Simmons, D.M., Swanson, L. and Rosenfeld, M.G. (1988) A tissue-specific transcription factor containing a homeodomain specifies a pituitary phenotype. Cell, 55, 519-529.
Jeong, J.H., Orvis, J., Kim, J.W., McMurtrey, C.P., Renne, R. and Dittmer, D.P. (2004) Regulation and auto-regulation of the promoter for the latency-associated nuclear antigen of Kaposi's sarcoma-associated herpesvirus. J Biol Chem, 279, 16822-16831.
Jorcyk, C.L., Watson, D.K., Mavrothalassitis, G.J. and Papas, T.S. (1991) The human ETS1 gene: genomic structure, promoter characterization and alternative splicing. Oncogene, 6, 523-532.
Kitaura, H., Galli, I., Taira, T., Iguchi-Ariga, S.M. and Ariga, H. (1991) Activation of c-myc promoter by c-myc protein in serum starved cells. FEBS Lett, 290, 147-152.
Konig, H., Ponta, H., Rahmsdorf, U., Buscher, M., Schonthal, A., Rahmsdorf, H.J. and Herrlich, P. (1989) Auto-regulation of fos: the dyad symmetry element as the major target of repression. Embo J, 8, 2559-2566.
Kume, K., Zylka, M.J., Sriram, S., Shearman, L.P., Weaver, D.R., Jin, X., Maywood, E.S., Hastings, M.H. and Reppert, S.M. (1999) mCRY1 and mCRY2 are essential components of the negative limb of the circadian clock feedback loop. Cell, 98, 193-205.
Laffitte, B.A., Joseph, S.B., Walczak, R., Pei, L., Wilpitz, D.C., Collins, J.L. and Tontonoz, P. (2001) Auto-regulation of the human liver X receptor alpha promoter. Mol Cell Biol, 21, 7558-7568.
Legraverend, C., Antonson, P., Flodby, P. and Xanthopoulos, K.G. (1993) High level activity of the mouse CCAAT/enhancer binding protein (C/EBP alpha) gene promoter involves auto-regulation and several ubiquitous transcription factors. Nucleic Acids Res, 21, 1735-1742.
Li, Y., Bolten, C., Bhat, B.G., Woodring-Dietz, J., Li, S., Prayaga, S.K., Xia, C. and Lala, D.S. (2002) Induction of human liver X receptor alpha gene expression via an auto-regulatory loop mechanism. Mol Endocrinol, 16, 506-514.
Lombardi, L., Ciana, P., Cappellini, C., Trecca, D., Guerrini, L., Migliazza, A., Maiolo, A.T. and Neri, A. (1995) Structural and functional characterization of the promoter regions of the NFKB2 gene. Nucleic Acids Res, 23, 2328-2336.
Mahatan, C.S., Kaestner, K.H., Geiman, D.E. and Yang, V.W. (1999) Characterization of the structure and regulation of the murine gene encoding gut-enriched Kruppel-like factor (Kruppel-like factor 4). Nucleic Acids Res, 27, 4562-4569.
Majerus, M.A., Bibollet-Ruche, F., Telliez, J.B., Wasylyk, B. and Bailleul, B. (1992) Serum, AP-1 and Ets-1 stimulate the human ets-1 promoter. Nucleic Acids Res, 20, 2699-2703.
Martinotti, S., Richman, A. and Hayday, A. (1988) Disruption of the putative c-myc auto-regulation mechanism in a human B cell line. Curr Top Microbiol Immunol, 141, 264-268.
Mechta, F.a.Y., M. (1994) In Angel, P.E.a.H., P.A. (ed.), The Fos and Jun Families of Transcription Factors. CRC Press, Boca Raton, p. 115.
Meyer, T.E., Waeber, G., Lin, J., Beckmann, W. and Habener, J.F. (1993) The promoter of the gene encoding 3',5'-cyclic adenosine monophosphate (cAMP) response element binding protein contains cAMP response elements: evidence for positive auto-regulation of gene transcription. Endocrinology,

132, 770-780.
Molina, C.A., Foulkes, N.S., Lalli, E. and Sassone-Corsi, P. (1993) Inducibility and negative auto-regulation of CREM: an alternative promoter directs the expression of ICER, an early response repressor. Cell, 75, 875-886.
Muralidharan, V., Sweet, T., Nadraga, Y., Amini, S. and Khalili, K. (2001) Regulation of Puralpha gene transcription: evidence for auto-regulation of Puralpha promoter. J Cell Physiol, 186, 406-413.
Narimatsu, M., Maeda, H., Itoh, S., Atsumi, T., Ohtani, T., Nishida, K., Itoh, M., Kamimura, D., Park, S.J., Mizuno, K., Miyazaki, J., Hibi, M., Ishihara, K., Nakajima, K. and Hirano, T. (2001) Tissue-specific auto-regulation of the stat3 gene and its role in interleukin-6-induced survival signals in T cells. Mol Cell Biol, 21, 6615-6625.
Neuman, E., Flemington, E.K., Sellers, W.R. and Kaelin, W.G., Jr. (1994) Transcription of the E2F-1 gene is rendered cell cycle dependent by E2F DNA-binding sites within its promoter. Mol Cell Biol, 14, 6607-6615.
Nicolaides, N.C., Gualdi, R., Casadevall, C., Manzella, L. and Calabretta, B. (1991) Positive auto-regulation of c-myb expression via Myb binding sites in the 5' flanking region of the human c-myb gene. Mol Cell Biol, 11, 6166-6176.
Niehof, M., Kubicka, S., Zender, L., Manns, M.P. and Trautwein, C. (2001) Auto-regulation enables different pathways to control CCAAT/enhancer binding protein beta (C/EBP beta) transcription. J Mol Biol, 309, 855-868.
Nishikawa, K., Kobayashi, M., Masumi, A., Lyons, S.E., Weinstein, B.M., Liu, P.P. and Yamamoto, M. (2003) Self-association of Gata1 enhances transcriptional activity in vivo in zebra fish embryos. Mol Cell Biol, 23, 8295-8305.
Oberkofler, H., Schraml, E., Krempler, F. and Patsch, W. (2003) Potentiation of liver X receptor transcriptional activity by peroxisome-proliferator-activated receptor gamma co-activator 1 alpha. Biochem J, 371, 89-96.
Oguchi, H. and Kimura, S. (1998) Multiple transcripts encoded by the thyroid-specific enhancer-binding protein (T/EBP)/thyroid-specific transcription factor-1 (TTF-1) gene: evidence of auto-regulation. Endocrinology, 139, 1999-2006.
Ohtani-Fujita, N., Fujita, T., Takahashi, R., Robbins, P.D., Dryja, T.P. and Sakai, T. (1994) A silencer element in the retinoblastoma tumor-suppressor gene. Oncogene, 9, 1703-1711.
Oka, T., Komuro, I., Shiojima, I., Hiroi, Y., Mizuno, T., Aikawa, R., Akazawa, H., Yamazaki, T. and Yazaki, Y. (1997) Auto-regulation of human cardiac homeobox gene CSX1: mediation by the enhancer element in the first intron. Heart Vessels, Suppl 12, 10-14.
Oka, T., Rairkar, A. and Chen, J.H. (1991) Structural and functional analysis of the regulatory sequences of the ets-1 gene. Oncogene, 6, 2077-2083.
Okumura-Nakanishi, S., Saito, M., Niwa, H. and Ishikawa, F. (2005) Oct-3/4 and Sox2 regulate Oct-3/4 gene in embryonic stem cells. J Biol Chem, 280, 5307-5317.
Okuno, Y., Huang, G., Rosenbauer, F., Evans, E.K., Radomska, H.S., Iwasaki, H., Akashi, K., Moreau-Gachelin, F., Li, Y., Zhang, P., Gottgens, B. and Tenen, D.G. (2005) Potential auto-regulation of transcription factor PU.1 by an upstream regulatory element. Mol Cell Biol, 25, 2832-2845.
Orkin, S.H. (1995) Hematopoiesis: how does it happen? Curr Opin Cell Biol, 7, 870-877.
Pankratova, E.V., Sytina, E.V., Luchina, N.N. and Krivega, I.V. (2003) The regulation of the Oct-1 gene transcription is mediated by two promoters. Immunol Lett, 88, 15-20.
Saadane, N., Alpert, L. and Chalifour, L.E. (2000) Altered molecular response to adrenoreceptor-induced cardiac hypertrophy in Egr-1-deficient mice. Am J Physiol Heart Circ Physiol, 278, H796-805.
Sadow, P.M., Chassande, O., Koo, E.K., Gauthier, K., Samarut, J., Xu, J., O'Malley, B.W. and Weiss, R.E. (2003) Regulation of expression of thyroid hormone receptor isoforms and coactivators in liver and heart by thyroid hormone. Mol Cell Endocrinol, 203, 65-75.
Saito, H., Yasumoto, K., Takeda, K., Takahashi, K., Fukuzaki, A., Orikasa, S. and Shibahara, S. (2002) Melanocyte-specific microphthalmia-associated transcription factor isoform activates its own gene promoter through physical interaction with lymphoid-enhancing factor 1. J Biol Chem, 277, 28787-28794.
Sassone-Corsi, P., Sisson, J.C. and Verma, I.M. (1988) Transcriptional auto-regulation of the proto-oncogene fos. Nature, 334, 314-319.
Schuettengruber, B., Simboeck, E., Khier, H. and Seiser, C. (2003) Auto-regulation of mouse histone deacetylase 1 expression. Mol Cell Biol, 23, 6993-7004.

Sears, R., Ohtani, K. and Nevins, J.R. (1997) Identification of positively and negatively acting elements regulating expression of the E2F2 gene in response to cell growth signals. Mol Cell Biol, 17, 5227-5235.
Shan, B., Chang, C.Y., Jones, D. and Lee, W.H. (1994) The transcription factor E2F-1 mediates the auto-regulation of RB gene expression. Mol Cell Biol, 14, 299-309.
Sivak, L.E., Tai, K.F., Smith, R.S., Dillon, P.A., Brodeur, G.M. and Carroll, W.L. (1997) Auto-regulation of the human N-myc oncogene is disrupted in amplified but not single-copy neuroblastoma cell lines. Oncogene, 15, 1937-1946.
Smith, J.E., 3rd and Cronmiller, C. (2001) The Drosophila daughterless gene auto-regulates and is controlled by both positive and negative cis regulation. Development, 128, 4705-4714.
Spencer, J.A. and Misra, R.P. (1996) Expression of the serum response factor gene is regulated by serum response factor binding sites. J Biol Chem, 271, 16535-16543.
Sun, H. and Taneja, R. (2000) Stra13 expression is associated with growth arrest and represses transcription through histone deacetylase (HDAC)-dependent and HDAC-independent mechanisms. Proc Natl Acad Sci U S A, 97, 4058-4063.
Tanabe, A., Kumahara, C., Osada, S., Nishihara, T. and Imagawa, M. (2000) Gene expression of CCAAT/enhancer-binding protein delta mediated by auto-regulation is repressed by related gene family proteins. Biol Pharm Bull, 23, 1424-1429.
Tapscott, S.J. and Weintraub, H. (1991) MyoD and the regulation of myogenesis by helix-loop-helix proteins. J Clin Invest, 87, 1133-1138.
Ten, R.M., Paya, C.V., Israel, N., Le Bail, O., Mattei, M.G., Virelizier, J.L., Kourilsky, P. and Israel, A. (1992) The characterization of the promoter of the gene encoding the p50 subunit of NF-kappa B indicates that it participates in its own regulation. Embo J, 11, 195-203.
Thayer, M.J., Tapscott, S.J., Davis, R.L., Wright, W.E., Lassar, A.B. and Weintraub, H. (1989) Positive auto-regulation of the myogenic determination gene MyoD1. Cell, 58, 241-248.
Theill, L.E. and Karin, M. (1993) Transcriptional control of GH expression and anterior pituitary development. Endocr Rev, 14, 670-689.
Timchenko, N., Wilson, D.R., Taylor, L.R., Abdelsayed, S., Wilde, M., Sawadogo, M. and Darlington, G.J. (1995) Auto-regulation of the human C/EBP alpha gene by stimulation of upstream stimulatory factor binding. Mol Cell Biol, 15, 1192-1202.
Tou, L., Quibria, N. and Alexander, J.M. (2003) Transcriptional regulation of the human Runx2/Cbfa1 gene promoter by bone morphogenetic protein-7. Mol Cell Endocrinol, 205, 121-129.
Trieu, M., Ma, A., Eng, S.R., Fedtsova, N. and Turner, E.E. (2003) Direct auto-regulation and gene dosage compensation by POU-domain transcription factor Brn3a. Development, 130, 111-121.
Tsai, S.F., Strauss, E. and Orkin, S.H. (1991) Functional analysis and in vivo footprinting implicate the erythroid transcription factor GATA-1 as a positive regulator of its own promoter. Genes Dev, 5, 919-931.
Ulven, S.M., Dalen, K.T., Gustafsson, J.A. and Nebb, H.I. (2004) Tissue-specific auto-regulation of the LXRalpha gene facilitates induction of apoE in mouse adipose tissue. J Lipid Res, 45, 2052-2062.
Walker, W.H., Fucci, L. and Habener, J.F. (1995) Expression of the gene encoding transcription factor cyclic adenosine 3',5'-monophosphate (cAMP) response element-binding protein (CREB): regulation by follicle-stimulating hormone-induced cAMP signaling in primary rat Sertoli cells. Endocrinology, 136, 3534-3545.
Wang, X., Li, Z., Naganuma, A. and Ye, B.H. (2002) Negative auto-regulation of BCL-6 is bypassed by genetic alterations in diffuse large B cell lymphomas. Proc Natl Acad Sci U S A, 99, 15018-15023.
Xu, F., Li, H. and Jin, T. (1999) Cell type-specific auto-regulation of the Caudal-related homeobox gene Cdx-2/3. J Biol Chem, 274, 34310-34316.
Yamaguchi, Y., Sawada, J., Yamada, M., Handa, H. and Azuma, N. (1997) Auto-regulation of Pax6 transcriptional activation by two distinct DNA-binding subdomains of the paired domain. Genes Cells, 2, 255-261.
Zhou, M. and Ouyang, W. (2003) The function role of GATA-3 in Th1 and Th2 differentiation. Immunol Res, 28, 25-37.
Zhu, L., Xie, E. and Chang, L.S. (1995) Differential roles of two tandem E2F sites in repression of the human p107 promoter by retinoblastoma and p107 proteins. Mol Cell Biol, 15, 3552-3562.

Table S6. Mammalian transcription factors that do not directly auto-regulate***

Regulator	Symbol	Reference
USF1	USF1	This work
YY1	YY1	Yao et al. 1998
SF1	NR5A1	Woodson et al. 1997
MR	NR3C2	Meyer et al. 1994
LEF1	LEF1	Vadlamudi et al. 2005
PAX3	PAX3	Borycki et al. 1999
TBX1	TBX1	Ataliotis et al. 2005

*** **References**

Borycki, A. G., Li, J., Jin, F., Emerson, C. P., and Epstein, J. A. (1999). Pax3 functions in cell survival and in pax7 regulation. Development 126, 1665-1674.

Meyer, A. S., and Schmidt, T. J. (1994). In contrast to glucocorticoid receptors, mineralocorticoid receptors are not auto-regulated in rat distal colon epithelia. Endocrinology 134, 1163-1172.

Vadlamudi, U., Espinoza, H. M., Ganga, M., Martin, D. M., Liu, X., Engelhardt, J. F., and Amendt, B. A. (2005). PITX2, beta-catenin and LEF-1 interact to synergistically regulate the LEF-1 promoter. J Cell Sci 118, 1129-1137.

Woodson, K. G., Crawford, P. A., Sadovsky, Y., and Milbrandt, J. (1997). Characterization of the promoter of SF-1, an orphan nuclear receptor required for adrenal and gonadal development. Mol Endocrinol 11, 117-126.

Yao, Y. L., Dupont, B. R., Ghosh, S., Fang, Y., Leach, R. J., and Seto, E. (1998). Cloning, chromosomal localization and promoter analysis of the human transcription factor YY1. Nucleic Acids Res 26, 3776-3783.

Comparison between Rada-Iglesias et al. ENCODE data and Odom et al. promoter data

Rada-Iglesias et al. *Human Mol. Genetics* (2005) reports ChIP-chip results using the NHGRI-designed ENCODE array design in human hepatomas with HNF4α, FOXA2, and USF1, all of which are profiled in primary human hepatocytes in this manuscript. Because similar transcription factors were targeted in both studies, we compared the results of these experiments. We note limited experimental overlap, characterized by differences in the tissues, antisera, the genomic regions targeted by each platform, and platform composition (i.e. PCR-product arrays versus 60-mer DNA oligonucleotide-based arrays). Similarities include identical profiled transcription factors and related tissue phenotypes (Table S7).

We identified the collection of bound regions which overlap between these two platforms and compared the data between these two reports (Materials and methods). Of the apolipoprotein cluster described in detail in Rada-Iglesias, we found qualitative agreement between the two datasets. The remaining interactions showed modest correspondence (Table S8). The binding sequences for each transcriptional regulator experiment are shown (Figure S2).

Table S7. Tissues, antisera, and array platforms used for ChIP experiments in Rada-Iglesias et al. 2005 and Odom et al. 2006.

	Rada-Iglesias et al. 2005	Odom et al. 2006
Tissues	HepG2 hepatoma	Primary human donor hepatocytes
Platform	ENCODE 30 MB arrays, PCR products	Agilent 10-array sets, 100 MB, 60-mer oligonucleotides
Genome Build	HG16	HG17
HNF4α antisera	ActiveMotif #39076 (discontinued)	Sc-8987 or sc-6556
USF1 antisera	Sc-229 and/or Sc-8983	AB/HLH230 (Cemines)
HNF3β antisera	Sc-6554*	Gift of R. Costa

Table S8. Binding event overlap between Rada-Iglesias et al. 2005 and Odom et al. 2006.

Transcriptional regulator	Regions bound on ENCODE arrays present on Agilent arrays	Binding events observed in common
USF1	11	0^+
HNF4α	30	12
FOXA2	17	4

+ Two binding events in Rada-Iglesias showed enrichment in corresponding regions in our data, but were slightly under the binding cutoff used for our data analysis.

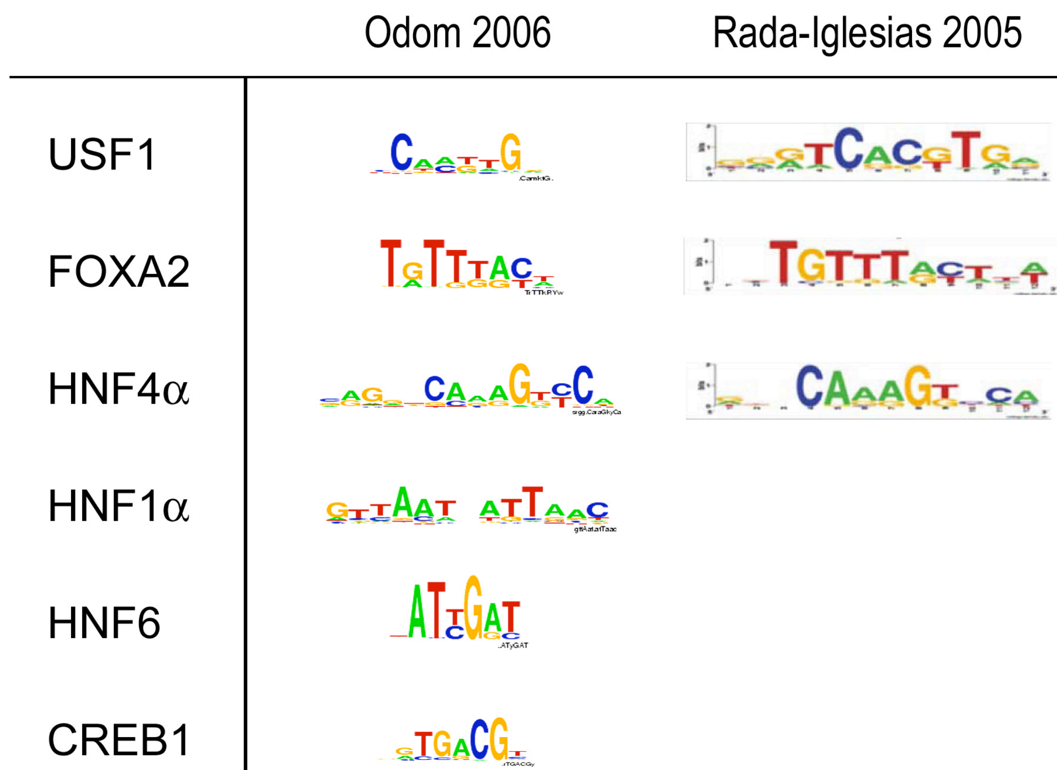


Figure S2. Binding sequences determined in Odom et al. 2006 and Rada-Iglesias et al. 2005. To identify motif for a factor, we use the motif discovery method THEME (see Materials and methods).

Table S9. Statistical significance of the sequence motifs reported here for six liver master regulators.

Transcriptional regulator	Cross-validation error	Significance of cross-validation errors (measured in standard deviations)++
FOXA2	0.36	7.73
HNF1α	0.35	6.59
HNF4α	0.31	7.49
HNF6	0.29	7.63
CREB1	0.22	15.42
USF1	0.43	2.42

++ Results were calculated using THEME (MacIsaac et al. 2006), as described in Materials and Methods. The significance of each motif was determined by comparing the observed cross-validation error (CV error) with the CV errors of motifs derived by applying THEME repeatedly to randomly selected input sequences, as described in MacIsaac et al. 2006. The significance of the CV error is reported as the number of standard deviations below the mean CV error observed in the randomized control calculations. The sequence motif discovered for USF1 is only marginally significant.

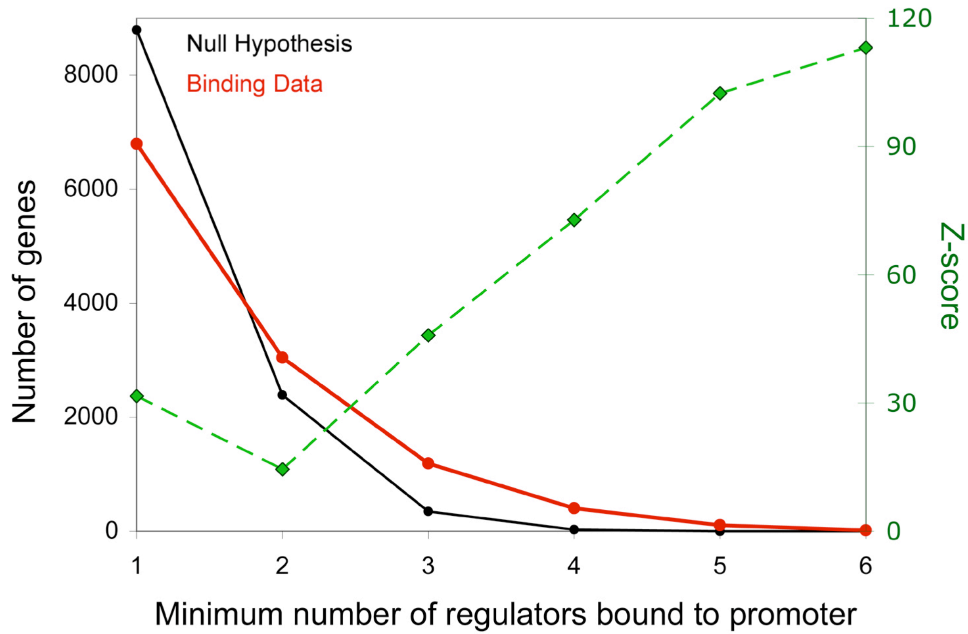


Figure S3. Combinatorial occupancy of promoter regions is significantly enriched, using expectations calculated assuming a binomial distribution of binding events. The expected (black) and actual (red) binding events are shown with the z-score of the difference (green).

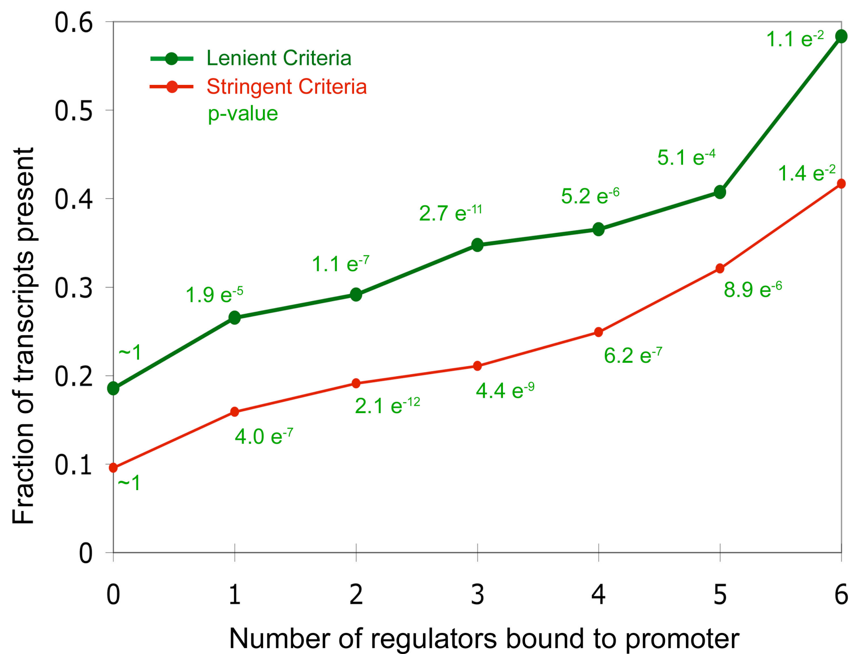


Figure S4. Number of bound transcription factors corresponds with gene expression in liver independent of stringency of analysis. The p-values were determined using a hypergeometric calculation, and transcripts present were established by analysis of publicly available compendiums of tissue-specific gene expression using lenient (green) and stringent (red) criteria (see methods below). The p-values for the higher number of transcription factors bound drop due to the decreasing sample size.

MATERIALS AND METHODS

Protocols can also be downloaded from http://web.wi.mit.edu/young/autoregulation/

Human tissues. Normal human hepatocytes were obtained through the Liver Tissue Procurement and Distribution System, Pittsburgh, Pennsylvania (funded by NIH Contract DK-92310). Cells were chemically cross-linked by the addition of a final concentration of 1% formaldehyde solution for 15 minutes at room temperature. Cells were rinsed twice with PBS or HBSS, frozen in liquid nitrogen, and stored at –80°C prior to use. The experiments reported here used hepatocyte mixtures created using healthy donor material obtained from multiple individuals, where the donor combinations were selected to maximize diversity of age and gender in liver mix (see supporting website for more information).

Regulator	Liver Used	Slide identifier format	Liver Used	Slide identifier format
HNF1α	Liver Mix 1	050828_HNF1A_hLiver_1_HP01a_slidecode	Liver Mix 2	050905_HNF1A_hLiver_2_HP01a_slidecode
HNF4α	Liver Mix 1	050826_HNF4A_hLiver_1_HP01a_slidecode	Liver Mix 2	050830_HNF4A_hLiver_2_HP01a_slidecode
FOXA2	Liver Mix 5	050920_FOXA2_hLiver_1_HP01a_slidecode	Liver Mix 2	050924_FOXA2_hLiver_2_HP01a_slidecode
HNF6	Liver Mix 1	050607_HNF6_hLiver_1_HP01a_slidecode	Liver Mix 2	050710_HNF6_hLiver_3_HP01a_slidecode
USF1	Liver Mix 3	050922_USF1_hLiver_1_HP01a_slidecode		
CREB1	Liver Mix 3	050926_pCREB_hLiver_1_HP01a_slidecode		

Chromatin immunoprecipitation protocol. Protocols describing materials and methods are available at http://web.wi.mit.edu/young/autoregulation/ and have been previously reported (Odom et al. 2004, Boyer et al. 2005). In short, frozen cells were resuspended, and lysed. Sonication conditions were determined for each cell batch to maximize chromatin sizes between 300-700 bp, and sonications were performed using a Misonix Sonicator 3000, typically at power 7 for ten x 30 second pulses (90 second pause between pulses) at 4°C in an ice-water bath. The lysate was clarified by centrifugation at 14,000 rpm at 4°C, and incubated 8 hours at 4°C with 100 μl of Dynal Protein G or Protein A magnetic beads and 10 μg of the appropriate antibody (Table S1). The purified DNA was blunted and ligated to unidirectional linker and amplified using a two-stage PCR protocol. Amplified DNA was labeled and purified using Invitrogen Bioprime random primer labeling kits. Labeled DNA was combined (5 – 6 μg each of immunoenriched Cy5 labeled and input Cy3 labeled DNA) and hybridized with tumbling to arrays in Agilent hybridization chambers for 40 hours at 40°C. Arrays were then washed and scanned per the manufacturer's instructions.

Microarray design. To more accurately identify the transcriptional regulatory circuitry in human hepatocytes, we have investigated genome-wide occupancy of transcriptional regulators at higher resolution than prior experimental designs used to explore liver-specific gene regulation. Previous designs have used self-printed DNA microarrays with one or two DNA elements to capture signals from each promoter region, covering up to 3 kb at approximately 3,000 transcription start sites, or 1 kb at approximately 13,000 transcription start sites (Friedman et al. 2004, Phuc Le et al. 2005, Rubins et al. 2005, Odom et al. 2004, Zhang et al. 2005). These arrays represented substantially less than whole genome coverage, and were either biased towards promoter regions identified as being present in cDNA libraries isolated from specific tissues (Friedman et al. 2004, Phuc Le et al. 2005, Rubins et al. 2005) or they represented promoters of earlier and less complete genomic annotations derived from NCBI databases (Odom et al. 2004, Zhang et al.

2005). A recent report (Rada-Iglesias et al. 2005) also uses self-print PCR-based ENCODE arrays (www.genome.gov) designed to capture a cross-section of genome features including the apolipoprotein cluster, combined with ChIPs against many of the same transcription factors we report. ENCODE arrays are however not principally gene-centric and have limited overlap with the data reported here (see above).

To overcome these limitations, we used commercial DNA microarrays with custom designs that represent a comprehensive collection of regulatory regions that control most predicted and essentially all known genes (Boyer et al. 2005). The results we obtain with this platform provide a more complete picture of transcription factor occupancy in human hepatocytes by key master regulators, and produce the first high-resolution maps of transcriptional regulation by a core set of master regulators in primary human hepatocytes.

A comprehensive collection of well-characterized transcription start sites was determined by identifying all transcription start sites described in five different databases: RefSeq, Ensembl, MGC, VEGA (www.vega.sanger.ac.uk) and Broad (www.broad.mit.edu). The first three are publicly available databases for gene annotation, the VEGA and Broad data sets are manually annotated databases. We identified all the transcription start sites present in any two of these five databases from the May 2004 build of the human genome.

We then designed DNA microarrays that contain 60-mer oligonucleotide probes covering the region from –8 kb to +2 kb relative to the transcript start sites for these 17,917 annotated human genes (Boyer et al. 2005). Although transcription factors are known to regulate genes from distances greater than 8 kb, most known binding sites for human transcription factors occur within 8 kb of the transcription start site. The sites occupied by transcription factors afforded peaks of ChIP-enriched DNA that span neighboring probes (examples Figure S1). These promoter regions averaged one 60-mer for each 250 bases of sequence, and allowed identification of in vivo binding sites to within tens of bases of position in the genome.

Identification of bound regions. To determine bound regions on the arrays, we developed an algorithm to incorporate information from neighboring probes (described in detail in Boyer et al. 2005). For each 60-mer, we calculated the average probability score of the 60-mer and its two immediate neighbors. If a feature was flagged as abnormal during scanning, we assigned it a neutral contribution to the average probability score. Similarly, if the nearest adjacent feature was beyond 1000 bp, we assigned a neutral contribution. The distance threshold of 1000 bp was determined based on the maximum size of labeled DNA fragments hybridized to the array using the shear distribution of the DNA amplified from the chromatin immunoprecipitation. The maximum fragment size ranged from 500 to 700 bp, depending on the particular experiment.

This set of averaged values gave us a distribution that was subsequently used to calculate p-values of average probe sets. The probe set p-value was determined for each factor based on a number of criteria, including capture of known positives and statistical measurements of the distribution of noise among control probes in the array data.

Comparison of gene expression and location data. Gene expression datasets used here were from the Novartis Gene Expression Atlas (http://wombat.gnf.org/) (Su et al. 2004). Expression probes were mapped to genes in the binding data using the knownToU133 and knownToGNF1 tables from UCSC (http://genome.ucsc.edu/). Careful cross-mapping of the genomic location of the complete set of probes in the ten-array set and the genes present on the expression arrays (both as described above) afforded a set of 13,243 genes where there was location and expression data that corresponded to the identical genes.

For the stringent analysis displayed in Figure 1B, we considered a gene present if and only if the expression is called as present (that is, 'P') in both replicates of human liver from the Novartis datasets. For genes with more than one expression probe, we further maximized stringency by requiring that all probes be uniformly present to call a gene expressed. Implementing these criteria identifies 1752 genes expressed in human liver.

Using this list of expressed genes, we then examined multi-input patterns (e.g. exactly 3 factors bound, etc.) to determine the fraction of genes within each multi-input motif expressed. A positive trend exists between the number of factors bound and the fraction of genes expressed, with p-values calculated using a hypergeometric distribution.

To confirm that this trend was not a result of the stringency of our criteria for defining a gene as expressed, we repeated the analysis using different stringencies. For comparison, Figure 1B compares our most lenient and strictest criteria for expression. The strict case is described above. For the lenient case, using a single present call among the probes that represent a gene to call the presence of a transcript, we identify 3149 genes as expressed. Combining these genes with the location data affords Figure S3, which incorporates the data displayed in Figure 1B for comparison. p-values for the expression-location comparisons for the higher number of genes bound (e.g. 6-factor multi-input motifs) drop due to the small sample sizes.

Identification of transcriptional regulatory motifs. We used algorithms previously reported (Lee et al. 2002) to determine the basic regulatory motifs described here. Briefly (partially paraphrased from Odom et al 2004), two data matrixes were created. The matrix consists of binary entries Dij, where we indicate a binding event of regulator j to intergenic region i as a 1; otherwise, as 0 is placed in this location. The regulator matrix R is a subset of D, containing only the rows corresponding to the intergenic region assigned to each regulator, in the same order as the columns of regulators.

Auto-regulatory motifs were identified as a non-zero entry on the diagonal of R. **Feedforward loops** were identified as a column of R with non-zero entries, which correspond to regulators bound. For each master regulator / secondary regulator pair, we then found all rows in the larger set of genes, D, bound by both regulators. **Multi-component loops** were identified as a column of R where the regulators to which it binds then binds the original regulator; this represents a multi-component loop of two. Repeat out to find larger loops. **Single input motifs** were identified as the collection of intergenic regions bound by only one regulator. **Multi-input motifs** were identified as the collections of intergenic regions bound by more than one regulator. Larger sized multi-input modules that employ more than 2 regulators could in principle be deconvoluted to combinations of smaller multi-input motifs. Notably, multi-input motifs are assembled from the largest size down, and smaller ones exclude genes previously identified as bound by more factors. For example, a gene like PCK1 is considered to be the target of a 6-input motif, and would not show up in smaller ones, though in principle any two regulators when taken independent of the complete dataset would form a 2-input motif at PCK1.

Identification and analysis of binding sequence motifs. Briefly, for each transcriptional regulator, we removed half the probes called as bound (determined as described above) that had the highest (worst) p-values (Lee et al. 2002, Boyer et al. 2005). We then sorted the remaining probes using alternative p-values that were calculated by incorporating data from adjacent probes, as described in Boyer et al. 2005. This ranking was used to identify the 100 best-bound probes, which were used as input sequences to the THEME program (MacIsaac et al. 2006) to determine a sequence

motif for each transcriptional regulator. Probe sequences stretched from the position of the probe 800 base pairs in each direction or half the distance to the neighboring probe, whichever was smaller.

THEME was employed to search for sequence motifs consistent with the family binding profiles corresponding to the PFAM family of each factor. We used with 3-fold cross-validation, and repeated the calculations three times with different random partitions of the data. We selected the family binding profile with the lowest median cross-validation error (CV error), and averaged the three motifs derived from this profile to use as our final motif for each regulator. To determine the statistical significance of the discovered sequence motifs, we compared the CV error of the reported motif with the CV errors of motifs derived by applying THEME to 50 sets of randomly selected input sequences using the same family binding profile and parameters, as described in MacIsaac et al. 2006.

We scanned all the bound probes with the reported motif to identify matches. To obtain the threshold for defining a match to the motifs, we re-ran the motifs through THEME using the same high-quality input sequences as before, but without the motif discovery step (the "-norefine" flag). This calculation determines the optimal threshold for classifying bound and not-bound regions using a support vector machine. The six supplemental files downloadable from the journal website are in THEME-output format and provide the coordinates of each match relative to the probe location.

Comparison of location data from Rada-Iglesias et al. 2005 to this work. The wiggle file reported in the supplemental materials of Rada-Iglesias et. al. 2005 was obtained and used as provided, with the recommended cutoff of 1.5 to identify binding events. We assigned each binding event to the closest gene within the hg16 genome build. We then determined how many of these binding events were contained within the region arrayed on our Agilent promoter arrays (described above), assuming localized equivalence between the hg16 and hg17 genome build (See Table S8: Regions bound on ENCODE arrays present on Agilent arrays). Finally, we determined how many of these genes were bound in our processed data using the standard error model and previously described cutoffs.

Data availability. Complete raw data from these experiments have been deposited into the public database ArrayExpress with accession number: **E-WMIT-9**

Materials and Methods References

Friedman, J. R., Larris, B., Le, P. P., Peiris, T. H., Arsenlis, A., Schug, J., Tobias, J. W., Kaestner, K. H., and Greenbaum, L. E. (2004). Orthogonal analysis of C/EBPbeta targets in vivo during liver proliferation. Proc Natl Acad Sci U S A 101, 12986-12991.

Odom, D. T., Zizlsperger, N., Gordon, D. B., Bell, G. W., Rinaldi, N. J., Murray, H. L., Volkert, T. L., Schreiber, J., Rolfe, P. A., Gifford, D. K., et al. (2004). Control of pancreas and liver gene expression by HNF transcription factors. Science 303, 1378-1381.

Boyer, L. A., Lee, T. I., Cole, M. F., Johnstone, S. E., Levine, S. S., Zucker, J. P., Guenther, M. G., Kumar, R. M., Murray, H. L., Jenner, R. G., et al. (2005). Core transcriptional regulatory circuitry in human embryonic stem cells. Cell 122, 947-956.

Phuc Le, P., Friedman, J. R., Schug, J., Brestelli, J. E., Parker, J. B., Bochkis, I. M., and Kaestner, K. H. (2005). Glucocorticoid receptor-dependent gene regulatory networks. PLoS Genet 1, e16.

Rada-Iglesias, A., Wallerman, O., Koch, C., Ameur, A., Enroth, S., Clelland, G., Wester, K., Wilcox, S., Dovey, O.M., Ellis, P.D., Wraight, V.L., James, K., Andrews, R., Langford, C., Dhami, P., Carter, N., Vetrie, D., Ponten, F., Komorowski, J., Dunham, I. and Wadelius, C. (2005) Binding sites for metabolic

disease related transcription factors inferred at base pair resolution by chromatin immunoprecipitation and genomic microarrays. Hum Mol Genet, 14, 3435-3447.
Rubins, N. E., Friedman, J. R., Le, P. P., Zhang, L., Brestelli, J., and Kaestner, K. H. (2005). Transcriptional networks in the liver: hepatocyte nuclear factor 6 function is largely independent of Foxa2. Mol Cell Biol 25, 7069-7077.
Su, A.I., Wiltshire, T., Batalov, S., Lapp, H., Ching, K.A., Block, D., Zhang, J., Soden, R., Hayakawa, M., Kreiman, G., Cooke, M.P., Walker, J.R. and Hogenesch, J.B. (2004) A gene atlas of the mouse and human protein-encoding transcriptomes. Proc Natl Acad Sci U S A, 101, 6062-6067.
Zhang, X., Odom, D. T., Koo, S. H., Conkright, M. D., Canettieri, G., Best, J., Chen, H., Jenner, R., Herbolsheimer, E., Jacobsen, E., et al. (2005). Genome-wide analysis of cAMP-response element binding protein occupancy, phosphorylation, and target gene activation in human tissues. Proc Natl Acad Sci U S A 102, 4459-4464.