DNA-binding specificity and in vivo targets of *Caenorhabditis elegans* nuclear factor I

Christina M. Whittle[a,1], Elena Lazakovitch[b,1], Richard M. Gronostajski[b,2], and Jason D. Lieb[a,2]

[a]Department of Biology, Carolina Center for the Genome Sciences, and Lineberger Comprehensive Cancer Center, University of North Carolina, Chapel Hill, NC 27599-3280; and [b]Department of Biochemistry, Program in Neurosciences Developmental Genomics Group, and Center of Excellence in Bioinformatics and Life Sciences, State University of New York at Buffalo, 701 Ellicott Street, Buffalo, NY 14203

Edited by Iva S. Greenwald, Columbia University, New York, NY, and approved May 21, 2009 (received for review December 18, 2008)

The conserved nuclear factor I (NFI) family of transcription factors is unique to animals and essential for mammalian development. The *Caenorhabditis elegans* genome encodes a single NFI family member, whereas vertebrate genomes encode 4 distinct NFI protein subtypes (A, B, C, and X). NFI-1-deficient worms exhibit abnormalities, including reduced lifespan, defects in movement and pharyngeal pumping, and delayed egg-laying. To explore the functional basis of these phenotypes, we sought to comprehensively identify NFI-1-bound loci in *C. elegans*. We first established NFI-1 DNA-binding specificity using an in vitro DNA-selection strategy. Analysis yielded a consensus motif of TTGGCA(N)$_3$TGCCAA, which occurs 586 times in the genome, a 100-fold higher frequency than expected. We next asked which sites were occupied by NFI-1 in vivo by performing chromatin immunoprecipitation of NFI-1 followed by microarray hybridization. Only 55 genomic locations were identified, an unexpectedly small target set. In vivo NFI-1 binding sites tend to be upstream of genes involved in core cellular processes, such as chromatin remodeling, mRNA splicing, and translation. Remarkably, 59 out of 70 (84%) of the *C. briggsae* orthologs of the identified targets contain conserved NFI binding sites in their promoters. These experiments provide a foundation for understanding how NFI-1 is recruited to unexpectedly few in vivo sites to perform its developmental functions, despite a vast over-representation of its binding motif.

ChIP-chip | protein-DNA interactions | transcription factor | binding site selection | cis regulatory elements

The Nuclear Factor I (NFI) DNA-binding proteins comprise a family of animal-specific transcription factors, with 4 members in vertebrates (NFIA, NFIB, NFIC, and NFIX) (1). NFI family members play major roles in specific aspects of mammalian organogenesis. Mice lacking either NFIA or NFIB exhibit forebrain defects and perinatal lethality (2, 3). In NFIB-deficient mice, these are accompanied by a failure in lung maturation (4). Mice lacking NFIC lack molar tooth roots and exhibit incisor defects (5). The human NFI proteins have analogous developmental functions. For example, individuals with 1 mutant copy of NFIA exhibit central nervous system malformations and urinary tract defects because of haploinsufficiency (3). The NFI family has a single member in simpler multicellular animals, including *Caenorhabditis elegans* and *Drosophila*. *C. elegans* NFI-1 (pronounced "NF one-one") has 4 known splice variants, each of which create very similar 86–88 kDa proteins. Amino acid homology to mammalian NFI proteins is limited to the DNA-binding domain. Loss of *nfi-1* in *C. elegans* results in a decreased rate of pharyngeal pumping, impaired egg-laying, altered motility, and a shortened lifespan (6). NFI-1 seems to function autonomously in the pharyngeal muscle with respect to pumping and lifespan (7).

Two primary questions surrounding NFI function relate to the identity of its binding targets in living cells, and how the selection of these in vivo targets is influenced by DNA sequence and chromatin. Previous studies of NFI-binding specificity used direct affinity isolation of transcription factor binding sites from the human genome (8), selection on oligonucleotide libraries (9), or bioinformatic comparison of motifs held in common by the promoters of putative NFI-regulated genes (10). NFI proteins have been linked to chromatin because both NFI and glucocorticoid receptor (GR) binding influence mouse mammary tumor virus (MMTV) promoter chromatin structure (11), and NFI proteins have functional and physical interactions with histones H1 and H3 (12, 13).

Here, we show that the binding specificity of NFI-1 is similar to the consensus sequence recognized by vertebrate NFI orthologs. Remarkably, NFI-1 motifs are present at ≈100 times their expected frequency in the worm genome. Through ChIP experiments, we show that despite this high abundance, only a small fraction of the sites appear to be occupied in vivo. These targets appear to be evolutionarily conserved because orthologs of *C. elegans* NFI-1 targets in *C. briggsae* contain conserved NFI motifs in their promoters. In addition, 2 of the mouse and human orthologs of the *C. elegans* NFI-1 targets contain conserved motifs in their promoters and may have relevance to NFI family members knockout phenotypes.

Results

The DNA Binding Specificity of *C. elegans* NFI-1 Is Nearly Identical to That of Vertebrate NFIs. We reasoned that unbiased identification of NFI-1 binding sites with naked genomic DNA in the absence of influence from chromatin and other cofactors would provide a foundation for investigation of the in vivo targets of NFI-1 [supporting information (SI) Fig. S1]. *C. elegans* genomic DNA was digested with Sau3A and fragments were selected for binding to immobilized GST-NFI-1 fusion protein. After elution, ligation of linkers, and PCR amplification, the selected DNA was subjected to 2 additional rounds of selection and amplification, following by TA-cloning and sequencing. The efficiency of each round of genomic selection was assessed by competition with a perfect-match oligonucleotide in a gel-shift assay (see *Materials and Methods*). As expected, DNA derived from each successive round of selection was increasingly able to compete with the perfect-match oligonucleotide (see Fig. S1). After 3 rounds of selection, 53 clones were obtained, representing 14 unique sequences. Alignment of the 14 sequences revealed a consensus sequence of TTGGCA(N)$_3$TGCCAA, which is nearly identical to the binding sequence reported previously for vertebrate NFIs [TTGGC-(N)$_5$GCCAA] (Fig. 1*A*) (see *Materials and Methods*) (10). The *C. elegans* consensus appears to have additional specificity at nucleotides 8 and 12, which flank the central spacer region (Fig. 1*B*).

Author contributions: C.M.W., E.L., R.M.G., and J.D.L. designed research; C.M.W. and E.L. performed research; C.M.W., E.L., R.M.G., and J.D.L. analyzed data; and C.M.W., E.L., R.M.G., and J.D.L. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

The data reported in this paper have been deposited in the NCBI Gene Expression Omnibus (GEO) database (www.ncbi.nlm.nih.gov/geo), accession no. GSE13918.

[1]C.M.W. and E.L. contributed equally to this work

[2]To whom correspondence may be addressed. E-mail: jlieb@bio.unc.edu or rgron@buffalo.edu.

This article contains supporting information online at www.pnas.org/cgi/content/full/0812894106/DCSupplemental.

GENETICS

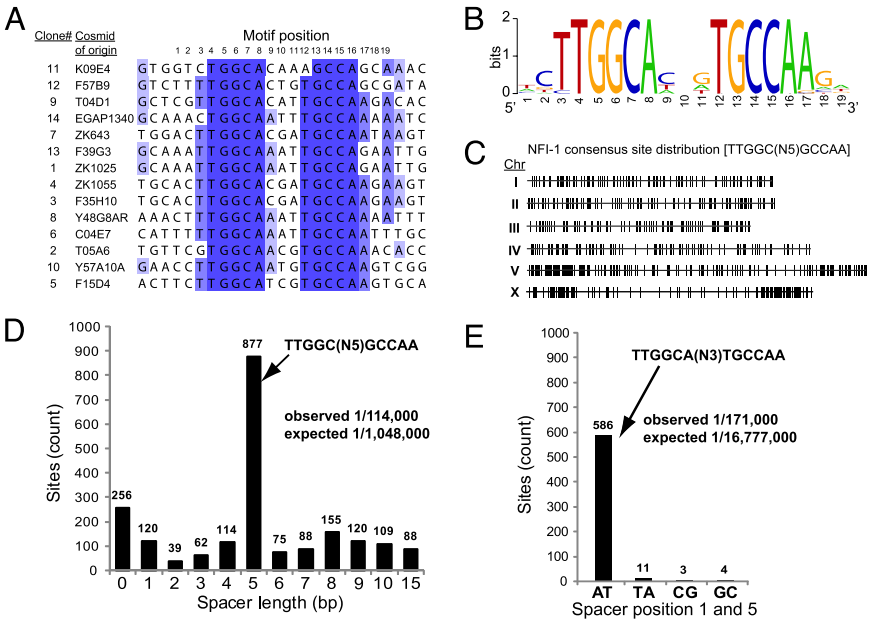

Fig. 1. The *C. elegans* NFI-1 in vitro consensus motif. (*A*) DNA from round 3 of in vitro selection was TOPO-cloned and 53 clones were sequenced. An alignment of the NFI-1 sites in the 14 unique sequences is shown, with shading indicating highly conserved positions. Dark blue indicates >90% identity, medium blue indicates 75 to 90% identity, and light blue indicates 50 to 75% identity. (*B*) The top-scoring motif discovered using MDscan (22) on the in vitro cloned sequences displayed as a sequence logo. (*C*) Genomic distribution of all 877 exact-match NFI-1 TTGGC(N)$_5$GCCAA consensus motifs in the *C. elegans* genome. Each vertical bar indicates one NFI-1 consensus motif. (*D*) Consensus NFI binding sites are over-represented, but not sites with different length spacer sequences. (*E*) The AT spacer motif is over-represented but not motifs containing TA, CG, or GC in positions 1 and 5 in the spacer region (equivalent to positions 8 and 12 in the 19-bp motif shown in B).

The DNA Sequence Bound by NFI-1 Occurs 100 Times More Often than Expected in the *C. elegans* Genome. We searched for the TTGGC(N)$_5$GCCAA motif in the *C. elegans* genome to create a catalog of potential NFI-1 binding sites. We found that NFI-1 consensus motifs containing either the N$_5$ spacer or the more specific AN$_3$T spacer were distributed relatively uniformly across each chromosome, with the exception of a noticeable clustering on the arms of chromosomes V and X (Fig. 1*C* and Fig. S2, and see *Discussion*). We were surprised to find 877 TTGGC(N)$_5$GCCAA motifs in the *C. elegans* genome, ≈10-fold more than would be predicted mathematically from genomic sequence composition (Fig. 1*D*). This over-representation was very specific to functional characteristics of the NFI-1 binding motif and was not observed in vertebrate genomes (Fig. S3). For example, over-representation was observed when the TTGGC and GCCAA portions of the motif were separated by a 5-nucleotide spacer, but not with 1-, 2-, 3-, 4-, 6-, 7-, 8-, 9-, 10-, or 15-nucleotide spacers (see Fig. 1*D*). Furthermore, the more specific TTGGCA(N)$_3$TGCCAA motif was present at over 100 times its expected frequency. Of the 877 TTGGC(N)$_5$GCCAA motifs throughout the genome, 67% (586) contain the A(N)$_3$T spacer, far above the expected frequency. No such over-representation was seen when TA, CG, or GC were used in the first and last positions of the 5-bp spacer (Fig. 1*E*), or when each half of the consensus is reversed (TGCCAA(N)$_3$TTGGCA).

Despite the Over-Representation of NFI-1-Binding Sequences, NFI-1 Binds to Few Targets in Vivo. To characterize direct in vivo targets of *C. elegans* NFI-1, we performed ChIP-chip experiments using extracts prepared from mixed stage wild-type worms (see *Materials and Methods*). Despite 877 perfect-match consensus sequences in the genome, we detected only 55 sites of NFI-1 binding (see *Materials and Methods*). To verify the microarray results, several loci were selected for analysis by qPCR: 9 ChIP-chip-positive loci with an NFI-1 motif, 4 ChIP-chip-positive loci without an NFI-1 motif, 6 ChIP-chip-negative loci with an NFI motif, and 4 randomly selected ChIP-chip-negative loci (Fig. S4). The qPCR results confirm the in vivo specificity of NFI-1 for the consensus motif and validate the microarray results. For example, at the *eft-2* promoter, which contains a single NFI-1 motif, the highest enrichment as measured by qPCR occurs directly at the NFI-1 motif (Fig. 2*A*).

The 55 NFI-1 target loci correspond to 85 downstream genes involved in basal cellular functions, including transcription, translation, biosynthesis, and GTPase signaling (Table 1). NFI-1 targets tend to be highly expressed and expressed throughout the

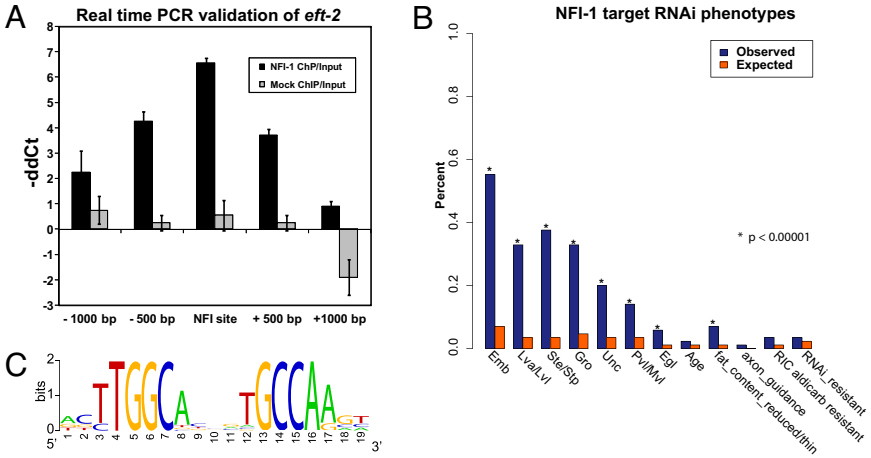

Fig. 2. The distribution of in vivo NFI-1 binding throughout the genome. (*A*) qPCR validation of NFI-1 binding at the *eft-2* promoter. ChIP, Input, and Mock DNA samples derived from independent ChIP biological replicates were analyzed with *ama-1* as a negative control. Bars on the graph represent corresponding DDCt (IP/Input) and their range. (*B*) RNAi phenotype frequency (wormbase release ws190) was plotted for the NFI-1 target genes (*blue*) compared to a randomly selected gene set of equal size (*orange*). (*C*) A sequence logo of the motif discovered from the in vivo binding targets of NFI-1 using MDscan (15).

Table 1. The direct in vivo targets of *C. elegans* NFI-1

Peak ID	Chr	Peak Center Coordinate	Peak *P*-value	Motif Count	Motif Distance to Target TSS	Gene Name (Systematic)	Gene Name (Common)	Gene Description
					Chromatin and Chromosome Structure			
NFI1_38	III	12368855	10^{-23}	1	−1046	*Y49E10.6*	*his-72*	Histone variant H3.3
NFI1_6	IV	4445469	10^{-43}	1	−79	*R08C7.3*	*htz-1*	Histone variant H2A.Z
NFI1_52	II	11979917	10^{-55}	1	−88	*Y17G7B.2*	*ash-2*	Ortholog of *Drosophila* Ash2, a member of the histone H3 (Lys4) methyltransferase complex
NFI1_16	I	2096862	10^{-19}	1	−277	*Y37E3.15*	*npp-13*	Nuclear pore complex protein
NFI1_40	III	13642931	10^{-51}	2	−27, 76	*T12D8.1*	*set-16*	SET (trithorax/polycomb) domain
NFI1_7	IV	4660215	10^{-44}	2	31, −288	*F29B9.2*		JmjC domain
					Transcription			
NFI1_5	IV	4024045	10^{-21}	2	2749, 2692	*W02C12.3*	*hlh-30*	Ortholog of human gene transcription factor binding to IGHM enhancer 3 (TFE3)
NFI1_51	II	11534541	10^{-29}	2	−927, −1074	*Y62F5A.1*	*mdt-8*	Homolog of mediator of RNA polymerase II transcription subunit 8
NFI1_46	II	6489336	10^{-19}	1	−191	*T28D9.2*	*rsp-5*	Splicing factor
NFI1_50	II	10658564	10^{-40}	1	−256	*D2089.1*	*rsp-7*	RNA binding protein
NFI1_32	III	8796617	10^{-18}	1	−1049	*F54F2.2*	*zfp-1*	Zinc finger protein homologous to human AF10
NFI1_30	III	4281786	10^{-65}	3	−1789, −1899, −2263	*R10E4.2*	*tag-310*	RNA binding motif protein
					Translation			
NFI1_19	I	9163212	10^{-31}	1	−42	*F25H5.4*	*eft-2*	Homolog of translation elongation factor 2 (EF-2)
NFI1_55	X	2242314	10^{-21}	1	−400	*F07D10.1*	*rpl-11.2*	Large ribosomal subunit L11 protein
NFI1_1	IV	654395	10^{-22}	1	−213	*K11H12.2*	*rpl-15*	Large ribosomal subunit L15 protein
NFI1_18	I	8815965	10^{-26}	1	−385	*F36A2.6*	*rps-15*	Small ribosomal subunit S15 protein
NFI1_43	V	10965223	10^{-43}	1	−13	*F17C11.9*		Elongation factor 1-gamma
					Cell Division			
NFI1_45	II	5593269	10^{-29}	2	299, −361	*C17G10.4*	*cdc-14*	Dual-specificity phosphatase homologous to yeast Cdc14p, required for cytokinesis
NFI1_17	I	2321403	10^{-45}	1	−140	*Y39G10AR.14*	*mcm-4*	MCM2/3/5 family with similarity to human MCM4
NFI1_25	III	1080156	10^{-39}	1	−153	*Y92C3B.1*	*kbp-4*	KNL (kinetochore null) binding protein
NFI1_37	III	12350884	10^{-23}	2	−144	*Y75B8A.30*	*pph-4.1*	Ser/Thr protein phosphatase
					GTPase signaling			
NFI1_13	IV	16946240	10^{-21}	0	na	*Y116A8C.12*	*arf-6*	Small GTP-binding protein of the ADP-ribosylation factor (ARF) family
NFI1_8	IV	10175885	10^{-18}	1	−273	*K04D7.1*	*rack-1*	Guanine nucleotide-binding protein
NFI1_33	III	9455738	10^{-42}	1	−255	*F54C8.5*	*rheb-1*	RHEB (Ras homology enriched in brain) GTPase Ortholog to mammalian Rheb1 GTPase
					Biosynthesis			
NFI1_20	I	10553443	10^{-47}	1	−51	*F25H2.5*		Nucleoside diphosphate kinase
NFI1_22	I	12651889	10^{-86}	1	−17	*H28O16.1*		ATP synthase alpha and beta subunits
					Other			
NFI1_2	IV	1709643	10^{-54}	1	−222	*K08D12.3*		Ortholog of human ZNF9 mutated in myotonic dystrophy type 2
NFI1_39	III	12549211	10^{-19}	2	−2122, −2202	*Y111B2A.8*		Ortholog of human gene AMP-activated protein kinase gamma subunit (PRKAG2)
NFI1_35	III	10747796	10^{-18}	1	−182	*K01G5.5*		Ortholog of human Dyskerin, which when mutated leads to X-linked dyskeratosis congenita
NFI1_24	III	349619	10^{-30}	0	na	*W07B3.2*	*gei-4*	Coiled-coil domain protein
NFI1_31	III	4899672	10^{-71}	2	222, 266	*F26F4.7*	*nhl-2*	NHL (ring finger b-box coiled coil) domain
NFI1_48	II	6588749	10^{-30}	1	−235	*C56C10.3*	*vps-32.1*	Related to yeast vacuolar protein sorting factor
NFI1_4	IV	2388879	10^{-59}	1	−354	*Y38F2AR.2*	*trap-3*	Translocon-associated complex (TRAP)
					Unknown			
NFI1_27	III	1925801	10^{-21}	1	na	*F53A3.6*		
NFI1_12	IV	16863426	10^{-19}	0	na	*T06A10.1*	*mel-46*	
NFI1_42	V	10715445	10^{-30}	2	−303, −748	*T28B11.1*		F-box protein
NFI1_3	IV	2219213	10^{-69}	2	−3221, −3077	*T04C4.1*		
NFI1_28	III	2375528	10^{-52}	1	−144	*H14E04.2*		
NFI1_49	II	9061539	10^{-49}	2	−802, −1381	*T24B8.3*		
NFI1_21	I	11610896	10^{-45}	1	−158	*C17D12.7*		
NFI1_34	III	10406926	10^{-38}	1	100	*M03C11.2*		
NFI1_54	II	14251210	10^{-35}	1	−776	*Y54G11A.2*		
NFI1_41	V	7727509	10^{-32}	3	−776, −406, −322	*Y97E10C.1*		
NFI1_15	I	1033830	10^{-31}	1	−129	*Y34D9A.3*		
NFI1_36	III	11935678	10^{-27}	1	−2519	*Y56A3A.36*		
NFI1_44	V	18078867	10^{-25}	1	−1999	*Y59A8B.10*		
NFI1_9	IV	11149119	10^{-24}	1	−297	*F01G4.6*		
NFI1_26	III	1418696	10^{-24}	1	−347	*Y82E9BR.3*		
NFI1_10	IV	16401729	10^{-24}	0	na	*Y65A5A.1*		
NFI1_47	II	6491510	10^{-23}	1	146	*T28D9.11*		
NFI1_53	II	13597222	10^{-21}	0	na	*Y48E1B.11*		
NFI1_11	IV	16655791	10^{-21}	1	97	*Y51H4A.15*		
NFI1_29	III	2948158	10^{-18}	2	−938, −864	*H05C05.1*		
NFI1_23	I	12903356	10^{-18}	1	382	*Y18D10A.16*		
NFI1_14	I	628957	10^{-18}	0	na	*Y65B4A.6*		

A list of the 55 NFI-1 *in vivo* targets. "Peak center coordinate" refers to the center of the maximum probe coordinate; after peak finding, all peaks were assigned to a 1,500-bp window centered on this coordinate for motif analysis. "Peak *P*-value" refers to the CHIPOTle (34) *P*-value. Motif distance and annotations are to the nearest TSS; in the case of bidirectional promoters only the closest gene is shown. Peaks were annotated using the maximum probe center to the nearest gene using CEAS (35). Gene descriptions are based on wormbase release ws192.

GENETICS

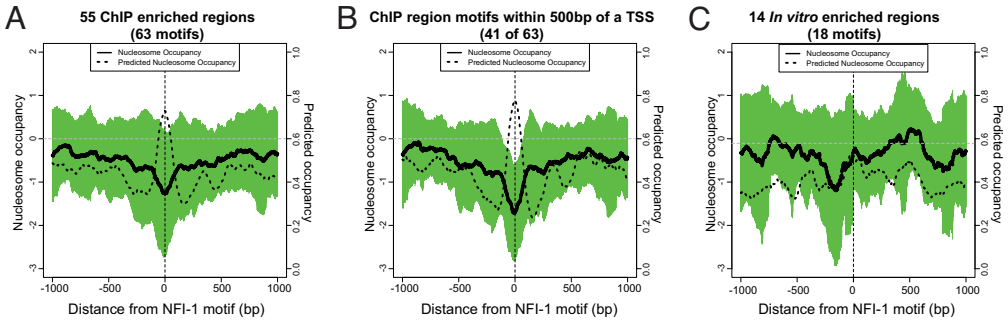

Fig. 3. Nucleosome occupancy surrounding NFI-1 binding sites. Mean nucleosome occupancy (*solid line*, right *y*-axis) (adjusted nucleosome stringency) and predicted nucleosome occupancy (*dashed line*, left *y*-axis) derived from published datasets (16, 17) were plotted across 2-kb windows centered on NFI-1 motifs found in the (*A*) in vivo targets, (*B*) in vivo targets within 500 bp of a TSS, or (*C*) in vitro targets. Green shading indicates ± 1 SD of in vivo occupancy; the vertical dotted line indicates the NFI-1 motif center.

C. elegans lifecycle (Fig. S5). All NFI-1-bound sites were located within 3.5 kb of a transcription start site (TSS) (Fig. S2*B*). NFI-1-bound loci were under-represented on chromosomes V and X, in contrast to the over-abundance of binding motifs on those chromosomes (see Fig. S2*A*).

Genes required for viability (Emb, Lva, Lvl phenotypes), fertility (Ste, Stp), growth (Gro), and vulval development (Pvl, Mvl) are over-represented in the NFI-1 target set (see Fig. 2*B*). In addition, we observe a significant over-representation of genes required for wild-type locomotion (Unc) and egg-laying (Egl). Misregulation of these target genes may contribute to the egg-laying and locomotion defects of *nfi-1* loss-of-function mutants. There is also a significant enrichment of genes that when mutated yield a "fat content reduced" phenotype (14), which may contribute to the lifespan phenotype of *nfi-1* mutants. The varied phenotypes associated with loss of function of NFI-1 targets are consistent with the varied, relatively weak phenotypes observed with *nfi-1* loss. Rather than functioning as a master regulator of any single process, NFI-1 may serve as a cofactor modulating transcription in several pathways.

The in Vivo DNA-Binding Specificity of NFI-1 Is Indistinguishable from Its Specificity in Vitro. We sought to determine whether the fact that NFI-1 bound few sites in vivo could be accounted for by a different DNA-binding specificity for NFI-1 in vivo. We ranked in vivo target loci by their amplitude, and extracted 1,500 bp of DNA sequence centered on the maximum probe for input to the motif discovery program MDscan (see *Materials and Methods*) (15). All 20 of the reported motifs were highly similar to the in vitro consensus sequence, TTGGC$(N)_5$CGGAA, including the 5-bp spacer and the preference for an A and T in positions 8 and 12, respectively. Of the 55 binding peaks, 48 (87%) contained at least one NFI-1 motif (see Table 1). The majority of peaks contained only one NFI-1 motif (see Table 1), although some contained up to 3 motifs as defined by the MDscan generated position weight matrix (Fig. 2*C*). We conclude that there is no functional difference between in vivo and in vitro NFI-1 DNA binding specificity.

The Paucity of in Vivo NFI-1 Targets Cannot Be Explained by Overly Stringent Criteria for Peak Definition. If our threshold for peak definition were too stringent, it would appear that NFI-1 was bound to few genomic targets, when in fact many were bound just under our threshold. If this were the case, the number of peaks detected should increase substantially as our threshold is lowered, and lowering the threshold should continue to capture loci with conserved NFI-1 consensus motifs. However, neither of these predictions holds. Instead, the number of peaks stays fairly constant over a wide range of cutoff values, and use of a lower stringency peak-finding cutoff substantially increases the number of peaks called that lack an NFI-1 motif (Fig. S6). For example, at the selected 1×10^{-12} peak cut-off, 87% of peaks contained an NFI-1 motif, compared to only 44% of peaks at a cut-off of 1×10^{-6}.

Sites Bound by NFI-1 in Vivo Exhibit Low Nucleosome Occupancy. We examined the relationship between nucleosome occupancy and NFI-1 binding throughout the genome. For the in vivo targets or the in vitro-selected sequences, computationally predicted (16) and in vivo mapped nucleosome occupancy (17) were plotted surrounding the NFI-1 motifs. At in vivo targets of NFI-1, the low nucleosome occupancy is centered at the NFI-1 motif, whereas high DNA-encoded nucleosome occupancy is predicted (Fig. 3*A*, Fig. S7). This observation is strongest at NFI-1 motifs located within the core promoter (less than 500 bp from the TSS) and is less prevalent at NFI motifs greater than 500 bp from the TSS (Fig. 3*B*). Within a 300-bp window surrounding the NFI-1 motif, the median local minimum of nucleosome occupancy occurs at + 6 bp from the motif at in vivo targets. In contrast, at the in vitro selected binding sites, measured nucleosome occupancy is no different from background at the NFI-1 motif, and DNA-encoded nucleosome occupancy is not predicted to be high at NFI-1 motifs (Fig. 3*C*). Thus, despite a strongly nucleosome-favoring DNA sequence at sites bound by NFI-1, nucleosomes are depleted from these sites, particularly at core promoters.

NFI-1 Target Genes Tend to Be Highly Conserved in *C. briggsae*, Mouse, and Human. Many of the *C. elegans* NFI-1 target genes are highly conserved across species (Fig. S8*A*, data not shown), leading us to examine motif conservation at the corresponding promoters in other animals. To find orthologs, for each *C. elegans* NFI-1 target we obtained the highest Blastp hit in *C. briggsae*, mouse, and human. We then obtained the 3-kb (*C. briggsae*) or 5-kb (mouse/human) promoter regions and searched for NFI-1 motifs (see *Materials and Methods*). Of the 78 *C. briggsae* orthologs of *C. elegans* targets, 84% contain promoter NFI-1 motifs, compared to 37% of *C. briggsae* orthologs of nontargets ($P = 9.7 \times 10^{-11}$), suggesting that *C. elegans* NFI-1 targets are functionally conserved in *C. briggsae* (Fig. S8*B*). Using a more stringent motif definition, 90% of *C. briggsae* orthologs of *C. elegans* targets have promoter NFI-1 motifs, compared to 24% of *C. briggsae* orthologs of nontargets ($P = 5.3 \times 10^{-24}$). Four *C. elegans* targets [*F25H2.5*, *nhl-2* (Fig. 4*A*), *pph-4.1*, and *rpt-6*] have NFI motifs in both the orthologous mouse and human promoter (see Fig. S8*A*). At *Trim2*/*TRIM2*, NFI motifs occur upstream of 3 out of 4 mouse alternative promoters (Fig. 4*B*) and both human alternative promoters (Fig. 4*C*). This motif occurrence in both the mouse and human ortholog promoters suggests that some *C. elegans* NFI-1 targets may be conserved in vertebrates (see Fig. S8*A*).

Discussion

We examined the in vitro affinity and in vivo binding of a *C. elegans* transcription factor, NFI-1. We find that NFI-1 binds 55 loci

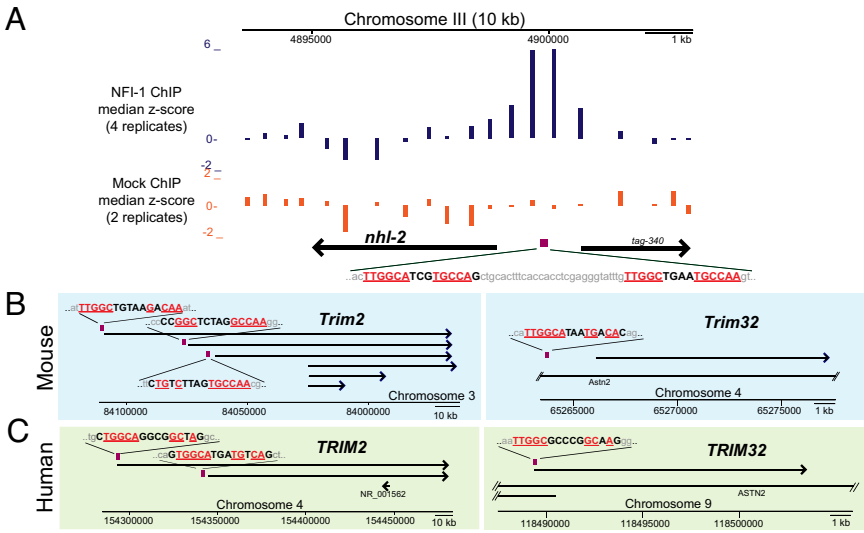

Fig. 4. Conservation of NFI-1 motifs in vertebrate orthologs of the *C. elegans* NFI-1 target *nhl-2*. The mouse and human ortholog of NFI-1 target gene *nhl-2*, *Trim2*, is highly expressed in neurons where deficiency leads to neurodegeneration (22). (*A*) A genome browser view showing NFI-1 ChIP (*blue*) or mock ChIP (*orange*) enrichment at the promoter of *nhl-2*. Motifs found at the *nhl-2* orthologs *Trim2* and *Trim32* within 5 kb of the mouse (*B*) or human (*C*) TSS are shown. In all panels, red underlined nucleotides represent perfect matches to the consensus motif; bolded letters represent the entire motif found; magenta boxes represent NFI motifs.

upstream of 85 genes involved predominantly in basal cellular functions, including transcription, translation, biosynthesis, and signaling. While very little functional information is known for over half of NFI-1 targets, a recent article linked NFI-1 target *rheb-1* to longevity, raising the possibility that the lifespan phenotype of *nfi-1* mutants may be mediated in part through *rheb-1* function (18). The identity of these targets provides insight regarding the types of targets that may be bound by vertebrate NFI orthologs. The majority of the *C. elegans* NFI-1 targets have well-conserved orthologs in more complex organisms (see Fig. 4, and Fig. S8). Deficiencies in mouse and human NFI proteins show brain and neurological defects, suggesting an important function in neuronal development (2, 3, 19, 20). Interestingly, 2 of the *C. elegans* targets (*nhl-2* and *F25H2.5*) whose mouse and human orthologs (Trim2/*Trim32* and *Nme2*, respectively) have promoter NFI motifs have been linked to neurological defects in mice (20–22). In addition, the promoters of *C. elegans nhl-2*, and both mouse and human homologs of *nhl-2* (*Trim2*/*Trim32*) contain NFI motifs (see Fig. 4). In *C. elegans*, NHI-2 is involved in miRNA regulation, and *Trim32* has been associated with myopathy in humans and defects in neuronal differentiation in mice (20). In addition, a second *nhl-2* ortholog, *Trim2*, has been associated with neurodegeneration in mice (22). Most recently, *F25H2.5*/*Nme2* has been shown to have similar regulation in both *C. elegans* and mouse in response to knockdown of FEH-1/Fe65, a protein involved neuronal development and implicated in the processing of aberrant proteins in Alzheimer's disease (21). These targets warrant further functional examination in both *C. elegans* and more complex metazoans, where the function of NFI may be conserved at promoters.

A remaining central question is why the NFI consensus sequence is specifically over-represented in the genome, while the number of targets bound in vivo is so small. NFI binding sites are found in of a number of viruses and retroviruses in mammals (23), raising the possibility that the worm motif TTGGCA$(N)_3$TGCCAA could be remnant of a virus or transposon. Indeed, the NFI motif is present in three *C. elegans* transposon-related repetitive elements [Ce000087, Ce000337, and Ce000357; wormbase release ws193) (24, 25)]. Detailed examination of the sequence context of these repeat-associated NFI-1 motifs may allow us to determine the biological significance of this unprecedented over-representation of a consensus transcription-factor binding motif.

Despite the specific nature of the motif over-representation, few stringent criteria for peak definition. An alternate technical explanation for the low number of targets in vivo is that because the ChIPs were performed from whole worms, we can detect only targets held in common between all tissues. While many of the NFI-1 target genes are expressed in several tissue types (primarily pharynx, intestine, and body-wall muscle), we identified other targets that appear to show specific expression in single-cell types, such as neurons (26). While not necessarily an indication of our ChIP, this suggests that we are able to detect tissue-specific NFI-1 targets with our approach. In addition, several loci containing NFI-1 motifs but not bound by NFI-1 in vivo were verified negatives by qPCR (see Fig. S4), suggesting a low false negative rate. Therefore, while technical explanations for the low number of targets cannot be ruled out, they seem unlikely.

The low number of in vivo NFI-1 targets, despite the motif over-representation and similarity to the in vitro NFI motif, suggests that additional factors play an important role in site selectivity in vivo. Our demonstration that in vivo-occupied NFI-1 sites tend to be centered within nucleosome-poor regions that are predicted on the basis of DNA sequence to be highly nucleosome occupied, suggests that NFI-1 is directed to targets as a consequence of low nucleosome occupancy introduced by other transacting factors. Alternatively, NFI-1 itself may actively displace nucleosomes. The best-studied chromatin model in which NFI proteins function is the MMTV promoter. Initial studies indicated that the nucleosome phasing was sequence-dependent (11), although more recent work has suggested that both NFI and glucocorticoid receptor (GR) binding may influence the MMTV-promoter chromatin structure (27). One widely supported model for hormone-induction of MMTV transcription is that NFI binding is normally excluded from the MMTV promoter by the presence of a sequence-dependent phased nucleosome (28), and that GR-recruited chromatin remodeling allows NFI to bind. Another model has been proposed, in which NFI binds with low affinity to the MMTV promoter and that synergism with GR binding allows for remodeling, high-affinity binding, and transcription activation (27). Taken together, the evidence suggests that multiple, perhaps synergistic, binding events with an unknown cofactor are likely required to direct NFI-1 to its targets and overcome high DNA-encoded nucleosome occupancy, but a definitive demonstration awaits further study.

Materials and Methods

In Vitro Genomic Selection. In vitro genomic selection was performed as previously described (29). *Sau3A*I-digested genomic DNA from a mixed-age population of N2 worms was incubated with NFI-1-GST Sepharose. DNA was eluted with glutathione and purified. DNA was ligated to linkers, PCR amplified, and sub-

GENETICS

jected to 2 additional rounds of selection. After the last round of selection and amplification, DNA was TOPO-cloned (Invitrogen) and all clones were sequenced.

DNA-Binding Assays. Gel mobility-shift assays and competition analysis were performed as described previously (6). Worm extracts were prepared from dounce-homogenized mixed-age worms in buffer described previously (6). Recombinant GST-NFI-1 protein (100 ng) and the labeled oligonucleotides containing a wild-type NFI binding site (wt) 5′AGGTCTGGCTTTGGGCCAAGAGCCGC or a site with a single-point mutation (mut) 5′AGGTCT**c**GCTTTGGGCCAAGAGCCGC shown previously to abolish the binding of vertebrate NFI proteins (30), were used. A 100-fold molar excess of unlabeled PCR amplified DNA from each round of selection was added to the indicated samples.

Chromatin Immunoprecipitation. Rabbit polyclonal antiserum was raised against recombinant NFI-1-GST fusion protein, described above. Antibody recognition of native NFI-1 protein bound to DNA was verified by gel mobility-shift assays (Fig. S9). N2 mixed-age worms were cross-linked as previously described (31). Cross-linked pellets (120–150 mg) were resuspended in ChIP lysis buffer (Upstate) sonicated using a Branson Sonifier 250 (output 30 and DutyCycle 30% setting) with 15 rounds of 10 1-sec pulses on ice. ChIP was performed using the ChIP assay kit (Upstate) with either 5-μl anti-NFI immune serum or preimmune serum with 5% set aside as the input sample. For ChIP-chip, samples were amplified by either ligation-mediated PCR (32) or a modified Whole Genome Amplification protocol [Sigma (33)], as previously described.

Microarrays and Data Extraction. DNA microarrays (Agilent Technologies) covering the entire *C. elegans* genome with 185,000 probes at an average start-to-start spacing of 600 bp were used for ChIP-chip (GEO Accession GPL7776). Four NFI-1 ChIP biological replicates and 2 NFI-1 preimmune ''mock'' ChIP-chip experiments were performed. Raw intensities for each ChIP were normalized by converting to standard Z-scores, and combined by taking the median of replicates. Raw and processed data can be accessed at NCBI GEO Accession GSE13918. Significant binding peaks were derived using a perl implementation of ChIPOTle (34). Peak maximum probes were extracted and annotated to the nearest gene using a *C. elegans* implementation of *Cis*-element annotation software (35) and hand-checked for accuracy (Wormbase release ws170).

ChIP-chip Data Analysis. A 1,500-bp window centered on the peak maximum probe was repeat masked using RepeatMasker (36). Peaks were ranked by maximum probe Z-score, and MDscan (15) was used for motif discovery. Matrixscan (32) was used to find motifs using the MDscan-generated position-weight matrix for the top-scoring motif. Distance to nearest TSS mappings, random window generation, and perfect-match motif finding were performed using custom Perl, Ruby, and R scripts (available upon request). Genome browser visualizations were obtained using the UCSC genome browser (http://genome.ucsc.edu), genome release ws170/ce4.

Modeling of nucleosome occupancy and MNase mapping of nucleosome occupancy and position (Adjusted Nucleosome Stringency) were derived from published datasets (16, 17). Raw expression data were obtained from the Stanford Microarray Database (http://smd.stanford.edu/) for a published *C. elegans* lifecycle time-course (37). Raw intensities for each expression microarray channel (mixed RNA reference or single stage) were percentile-ranked as an unbiased measure of relative RNA abundance.

Precomputed wormbase blastp hits were used to find orthologs of the *C. elegans* NFI-1 targets. The lowest e-value hit was chosen and 3 kb (*C. briggsae*) or 5 kb (mouse/human) upstream of the TSS was analyzed for motifs using Matrixscan (32) using the *C. elegans*-derived position-weight matrix. *C. briggsae* sequences and annotations were obtained for wormbase genome release ws190; mouse (Ensembl50/NCBI m37), and human (Ensembl50/NCBI36) sequences and annotations were obtained via Ensembl (http://www.ensembl.org/) and the UCSC genome browser.

qPCR Analysis of ChIP-chip Data. qPCR was used to determine relative amount of specific loci in IP, Input, and Mock (Preimmune) samples. qPCR was performed using iQ SYBR Green Supermix (Bio-Rad) on a Bio-Rad iCycler. One microliter of ChIP DNA or a 1:1,000 Dilution of input DNA was used in duplicate reactions. A locus negative for NFI-1 binding (*ama-1*) was used as an internal control to normalized quantification in qPCR reactions. Data are expressed as IP/Input where $DDCT = (Ct_{IP\_locusX} - Ct_{IP\_ama\text{-}1}) - (Ct_{Input\_locusX} - Ct_{Input\_ama\text{-}1})$.

ACKNOWLEDGMENTS. We thank Joanna Mieczkowska for assistance with experiments, X. Shirley Liu for the Matrixscan and *Cis*-element annotation software, and Yaniv Lubling and Eran Segal for providing the predictions of nucleosome occupancy. This work is supported by National Heart Lung and Blood Institute Grant 5R01HL080624–02 (to R.M.G.), a V Foundation for Cancer Research Scholar award (to J.D.L.), and in part by National Institutes of Health Training Grant T32 HD046369 (to C.M.W.).

1. Gronostajski RM (2000) Roles of the NFI/CTF gene family in transcription and development. *Gene* 249:31–45.
2. Steele-Perkins G, et al. (2005) The transcription factor gene Nfib is essential for both lung maturation and brain development. *Mol Cell Biol* 25:685–698.
3. das Neves L, et al. (1999) Disruption of the murine nuclear factor I-A gene (Nfia) results in perinatal lethality, hydrocephalus, and agenesis of the corpus callosum. *Proc Natl Acad Sci USA* 96:11946–11951.
4. Grunder A, et al. (2002) Nuclear factor I-B (Nfib) deficient mice have severe lung hypoplasia. *Mech Dev* 112:69–77.
5. Muse GW, et al. (2007) RNA polymerase is poised for activation across the genome. *Nat Genet* 39:1507–1511.
6. Lazakovitch E, et al. (2005) nfi-I affects behavior and life-span in *C. elegans* but is not essential for DNA replication or survival. *BMC Dev Biol* 5:24.
7. Lazakovitch E, Kalb JM, Gronostajski RM (2008) Lifespan extension and increased pumping rate accompany pharyngeal muscle-specific expression of nfi-1 in *C. elegans*. *Dev Dyn* 237:2100–2107.
8. Gronostajski RM, Nagata K, Hurwitz J (1984) Isolation of human DNA sequences that bind to nuclear factor I, a host protein involved in adenovirus DNA replication. *Proc Natl Acad Sci USA* 81:4013–4017.
9. Gronostajski RM (1986) Analysis of nuclear factor I binding to DNA using degenerate oligonucleotides. *Nucl Acids Res* 14:9117–9132.
10. Roulet E, et al. (2000) Experimental analysis and computer prediction of CTF/NFI transcription factor DNA binding sites. *J Mol Biol* 297:833–848.
11. Beato M (1996) Chromatin structure and the regulation of gene expression: remodeling at the MMTV promoter. *J Mol Med* 74:711–724.
12. Dusserre Y, Mermod N (1992) Purified cofactors and histone H1 mediate transcriptional regulation by CTF/NF-I. *Mol Cell Biol* 12:5228–5237.
13. Muller K, Mermod N (2000) The histone-interacting domain of nuclear factor I activates simian virus 40 DNA replication in vivo. *J Biol Chem* 275:1645–1650.
14. Ashrafi K, et al. (2003) Genome-wide RNAi analysis of *Caenorhabditis elegans* fat regulatory genes. *Nature* 421:268–272.
15. Liu XS, Brutlag DL, Liu JS (2002) An algorithm for finding protein-DNA binding sites with applications to chromatin-immunoprecipitation microarray experiments. *Nat Biotechnol* 20:835–839.
16. Kaplan N, et al. (2008) The DNA-encoded nucleosome organization of a eukaryotic genome. *Nature* 458:362–366.
17. Valouev A, et al. (2008) A high-resolution, nucleosome position map of *C. elegans* reveals a lack of universal sequence-dictated positioning. *Genome Res* 18:1051–1063.
18. Honjoh S, Yamamoto T, Uno M, Nishida E (2009) Signalling through RHEB-1 mediates intermittent fasting-induced longevity in *C. elegans*. *Nature* 457:726–730.
19. Campbell CE, et al. (2008) The transcription factor Nfix is essential for normal brain development. *BMC Dev Biol* 8:52.
20. Schwamborn JC, Berezikov E, Knoblich JA (2009) The TRIM-NHL protein TRIM32 activates microRNAs and prevents self-renewal in mouse neural progenitors. *Cell* 136:913–925.
21. Napolitano F, et al. (2008) A differential proteomic approach reveals an evolutionary conserved regulation of Nme proteins by Fe65 in *C. elegans* and mouse. *Neurochem Res* 33:2547–2555.
22. Balastik M, et al. (2008) Deficiency in ubiquitin ligase TRIM2 causes accumulation of neurofilament light chain and neurodegeneration. *Proc Natl Acad Sci USA* 105:12016–12021.
23. Nowock J, Borgmeyer U, Puschel AW, Rupp RAW, Sippel AE (1985) The TGGCA protein binds to the MMTV-LTR, the adenovirus origin of replication, and the BK virus enhancer. *Nucl Acids Res* 13:2045–2061.
24. Oosumi T, Garlick B, Belknap WR (1996) Identification of putative nonautonomous transposable elements associated with several transposon families in *Caenorhabditis elegans*. *J Mol Evol* 43:11–18.
25. Jurka J, et al. (2005) Repbase Update, a database of eukaryotic repetitive elements. *Cytogenet Genome Res* 110:462–467.
26. Hunt-Newbury R, et al. (2007) High-throughput in vivo analysis of gene expression in *Caenorhabditis elegans*. *PLoS Biol* 5:e237.
27. Hebbar PB, Archer TK (2007) Chromatin-dependent cooperativity between site-specific transcription factors in vivo. *J Biol Chem* 282:8284–8291.
28. Chavez S, Beato M (1997) Nucleosome-mediated synergism between transcription factors on the mouse mammary tumor virus promoter. *Proc Natl Acad Sci USA* 94:2885–2890.
29. Shostak Y, Van Gilst MR, Antebi A, Yamamoto KR (2004) Identification of *C. elegans* DAF-12-binding sites, response elements, and target genes. *Genes Dev* 18:2529–2544.
30. Goyal N, Knox J, Gronostajski R (1990) Analysis of multiple forms of nuclear factor I in human and murine cell lines. *Mol Cell Biol* 10:1041–1048.
31. Oh SW, et al. (2006) Identification of direct DAF-16 targets controlling longevity, metabolism and diapause by chromatin immunoprecipitation. *Nat Genet* 38:251–257.
32. Ercan S, et al. (2007) X chromosome repression by localization of the *C. elegans* dosage compensation machinery to sites of transcription initiation. *Nat Genet* 39:403–408.
33. O'Geen H, Nicolet CM, Blahnik K, Green R, Farnham PJ (2006) Comparison of sample preparation methods for ChIP-chip assays. *Biotechniques* 41:577–580.
34. Buck MJ, Nobel AB, Lieb JD (2005) ChIPOTle: a user-friendly tool for the analysis of ChIP-chip data. *Genome Biol* 6:R97.
35. Ji X, Li W, Song J, Wei L, Liu XS (2006) CEAS: *cis*-regulatory element annotation system. *Nucleic Acids Res* 34(Web Server issue):W551–W554.
36. Chen N (2004) *Current Protocols in Bioinformatics* (Wiley, Hoboken, NJ), Chapter 4, Unit 4-10, pp 4.10.1–4.10.14.
37. Jiang M ,et al. (2001) Genome-wide analysis of developmental and sex-regulated gene expression profiles in *Caenorhabditis elegans*. *Proc Natl Acad Sci USA* 98:218–223.