THE JOURNAL OF BIOLOGICAL CHEMISTRY VOL. 286, NO. 21, pp. 18641–18649, May 27, 2011

Genomic Selection Identifies Vertebrate Transcription Factor Fezf2 Binding Sites and Target Genes*[S]

Received for publication, March 2, 2011, and in revised form, April 4, 2011 Published, JBC Papers in Press, April 6, 2011, DOI 10.1074/jbc.M111.236471

Lishan Chen[‡], Jiashun Zheng[§], Nan Yang[‡1], Hao Li[§], and Su Guo[‡2]

From the [‡]Department of Bioengineering and Therapeutic Sciences, Programs in Human Genetics and Biological Sciences, and the [§]Department of Biochemistry and Biophysics and the California Institute for Quantitative Biomedical Research, University of California, San Francisco, California 94158

Identification of transcription factor targets is critical to understanding gene regulatory networks. Here, we uncover transcription factor binding sites and target genes employing systematic evolution of ligands by exponential enrichment (SELEX). Instead of selecting randomly synthesized DNA oligonucleotides as in most SELEX studies, we utilized zebrafish genomic DNA to isolate fragments bound by Fezf2, an evolutionarily conserved gene critical for vertebrate forebrain development. This is, to our knowledge, the first time that SELEX is applied to a vertebrate genome. Computational analysis of bound genomic fragments predicted a core consensus binding site, which identified response elements that mediated Fezf2-dependent transcription both *in vitro* and *in vivo*. Fezf2-bound fragments were enriched for conserved sequences. Surprisingly, ~20% of these fragments overlapped well annotated protein-coding exons. Through loss of function, gain of function, and chromatin immunoprecipitation, we further identified and validated *eomesa/tbr2* and *lhx2b* as biologically relevant target genes of Fezf2. Mutations in *eomesa/tbr2* cause microcephaly in humans, whereas *lhx2b* is a critical regulator of cell fate and axonal targeting in the developing forebrain. These results demonstrate the feasibility of employing genomic SELEX to identify vertebrate transcription factor binding sites and target genes and reveal Fezf2 as a transcription activator and a candidate for evaluation in human microcephaly.

Understanding the function of genomes is one of the most challenging tasks that biologists face in the post-genomic sequencing era. This requires not only the elucidation of the function of an individual gene, but also the characterization of regulatory gene networks. A critical entry point to the characterization of regulatory networks is the identification of transcription factor (TF)[3] binding sites and target genes.

Multiple technologies have been developed to define binding sites for TFs with unknown DNA binding profiles. These include chromatin immunoprecipitation (ChIP)-based methods (*e.g.* ChIP on a chip or ChIP-Seq) (1, 2), which uncover DNA fragments occupied by the factor either directly or indirectly *in vivo*. Because of the indirectly bound DNA fragments, ChIP methods are not as powerful in discovering DNA binding sites for the protein of interest. Additionally, ChIP methods require a ChIP-quality antibody and cells or tissue types in which the regulatory factor is abundantly expressed and active. In contrast, *in vitro* methods for defining TF binding sites identify directly bound DNA and can be used without knowledge of the conditions under which TFs are active. One well established *in vitro* method for identifying TF binding sites is based on systematic evolution of ligands by exponential enrichment (SELEX), which allows extraction of oligomers from an initially random pool of oligonucleotides based on their binding affinity for the transcription factor of interest (3–7). Although SELEX is highly effective in recovering DNA sequences that interact with the transcription factor of interest *in vitro*, it remains a difficult task to uncover target genes using the DNA motifs identified through SELEX because they are often short, highly degenerate, and thus can be found in numerous locations throughout the genome. This makes it difficult to prioritize efforts for target validation. To circumvent this problem, SELEX applications employing genomic DNA fragments have been reported in *Escherichia coli* (8) and *Caenorhabditis elegans* (9) that have successfully defined TF binding sites as well as identified *bona fide* target genes. However, it is not clear whether genomic SELEX would be successful when extended to the more complex vertebrate genome, which is roughly 20 times larger than that of *C. elegans*.

In this study, we employed *in vitro* genomic SELEX strategy to isolate zebrafish DNA fragments bound by the forebrain embryonic zinc finger 2 (Fezf2, also known as Fezl, Znf312, or Zfp312). Fezf2 is an evolutionarily conserved protein originally discovered for its selective expression in the developing forebrain (10, 11). Its activity is required for proper development of dopaminergic and serotonergic neurons and for regional patterning in the developing zebrafish forebrain (12–14). Fezf2 expression is also detected in the dorsal telencephalic adult progenitors that completely overlap with markers of neural stem cells, suggesting a potential role in regulating adult neurogenesis (15). In mice, Fezf2 is critical for the specification of cortical neurons, including the corticospinal motor neurons, which degenerate in patients suffering from amyotrophic lateral sclerosis, and also plays an important role in axonal outgrowth or

* This work was supported, in whole or in part, by National Institutes of Health Grant NS042626 (to S. G.).
[S] The on-line version of this article (available at http://www.jbc.org) contains supplemental Experimental Procedures, additional references, and Tables S1–S4.
[1] Present address: Dept. of Pathology, Stanford University, Stanford, CA 94305.
[2] To whom correspondence should be addressed. E-mail: su.guo@ucsf.edu.
[3] The abbreviations used are: TF, transcription factor; Fezf-2, forebrain embryonic zinc finger 2; hpf, hours post-fertilization; SELEX, systematic evolution of ligands by exponential enrichment; ZF, zinc finger.

targeting (16–19). Together, vertebrate Fezf2 performs evolutionarily conserved functions in forebrain development (20).

Fezf2 contains six C_2H_2-type zinc fingers located at its C terminus, likely functioning as a DNA-binding protein. Through *in vitro* genomic SELEX, we isolated zebrafish DNA fragments bound by Fezf2. Computational analyses uncovered a consensus core Fezf2 DNA binding site, which subsequently led to the identification of response elements in the selected genomic fragments. Biochemical binding, cell culture-based reporter assay as well as *in vivo* analyses in zebrafish embryos demonstrated that the identified response elements interacted with Fezf2 both *in vitro* and *in vivo*. Fezf2-bound genomic fragments were significantly enriched for conserved elements. Surprisingly, ~20% of these fragments overlapped protein-coding exons. Last but not the least, we revealed that the evolutionarily conserved fate determination genes, *eomesa/tbr2* and *lhx2b*, are direct targets of Fezf2 in regulating forebrain development.

EXPERIMENTAL PROCEDURES

Zebrafish Strains and Maintenance—Zebrafish (wild type and the *tof* mutant) were maintained and bred following standard procedures (40).

In Vitro Genomic Selection—GST-Fezf2 zinc finger (ZF) domain (amino acids 253–438) fusion protein was immobilized and incubated with 20 μg of zebrafish genomic DNA digested with Sau3A1 restriction endonuclease. Protein-DNA bead complexes were washed, and GST-ZF-bound DNA was eluted. After adapter ligation, eluted DNA was used for first PCR. First round PCR products were used in second round of binding/elution. After the second, third, or fourth rounds of PCR/binding/elution, purified PCR DNA products were digested with NotI restriction endonuclease (NEB) and subcloned into the NotI site of pBluescript II KS^+ vector (Stratagene) for sequencing. See also supplemental Experimental Procedures.

Computational Prediction of Fezf2 Core Consensus Binding Site and Analyses of Unique Fezf2-bound Selected Genomic Fragments—Several motif-finding algorithms, including BioProspector, AlignACE, and MEME, were used to identify the sequence motifs that are enriched in the unique fragments from the fourth round, using all genomic sequences as background. For details, see supplemental Experimental Procedures.

For the analyses of unique Fezf2-bound selected genomic fragments, we first determined whether one SELEX fragment contain the core binding motif by calculating the matrix similarity score using the derived Fezf2 positional weight matrix from BioProspector. Motifs with scores larger than the cutoff (0.8469) were regarded as a match to the core binding site of Fezf2. This cutoff is the smallest Matrix similarity score of the motifs on the SELEX fragments that were sequenced multiple times.

The six-way conservation scores considering the alignment across six species (zebrafish, frog, stickleback, tetraodon, human, and mouse) were downloaded from the UCSC genome browser (41, 42). Using their criteria, about 15% of the zebrafish genome are considered as conserved (*i.e.* can obtain good alignment). SELEX fragments overlapped with the conserved regions defined by these conservation scores were defined as conserved fragments. The background frequency of obtained conserved fragments were estimated by randomly drawing a set of none-repeat fragments with the same size distribution as the SELEX fragments. The enrichment p value of the conserved unique fragments was than calculated using binomial distribution. The exon and gene position information was obtained from the Ensembl database (Zv8; see supplemental Table S4).

Fluorescence Anisotropy—Fluorescein-labeled double-stranded oligonucleotides were used for fluorescence anisotropy measurements with GST-Fezf2 ZF. Excitation was performed at 485-nm wavelength, and emission was monitored at 515 nm using a Tecan fluorometer. Anisotropy calculations and K_d determinations by curve fitting were performed as described previously (9).

Cell Culture, Transient Transfection, and Luciferase Assays—HEK293 cells were transfected with pCS2 with or without Fezf2 cDNA, pGL4.10[*luc2*] plasmid containing three copies of the Fezf2 binding site (either WT or mutant versions), and pGL4.74[*hRluc*/TK] plasmid. Cell lysates were assayed for firefly and *Renilla* luciferase activities in Tecan luminometer (Tecan Group Ltd.). Reactions were carried out using the Dual-Luciferase Reporter Assay system (Promega). Data are reported as the mean ± S.D. of at least three independent experiments.

In Vivo Reporter Assay—WT or the mutant *fezf2* response elements were subcloned upstream of a basal E1b promoter driving *EGFP* expression in the pT2KXIGQ vector (15) and microinjected into one-cell embryos. At least 50 embryos were injected for each construct, and only embryos with normal morphology were considered for subsequent analysis. To block the Fezf2 protein activity, *fezf2* morpholino antisense oligonucleotides (14) were co-injected with reporter constructs. Live zebrafish embryos were photographed using a Zeiss Axioskop 2 plus epifluorescent compound microscope connected with a CCD camera.

Preparation and Test of an Antibody to Fezf2—A rabbit anti-Fezf2 antibody was made against the N-terminal amino acids 166–255 of Fezf2 fused with GST. The affinity-purified antibody was tested on zebrafish embryonic extracts, and the HEK293 cell lysates were transfected with or without Fezf2 through Western blotting analysis.

ChIP—ChIP was carried out as described previously (43, 44). In brief, embryonic lysates were immunoprecipitated with the affinity-purified rabbit anti-Fezf2 antibody or the preimmune serum. Chromatin DNA was purified by phenol extraction followed by ethanol precipitation and used for PCR analysis.

In Situ Hybridization and Immunostaining—RNA *in situ* hybridization and immunohistochemistry were performed as described (45).

RESULTS

Genomic SELEX Isolates DNA fragments That Bind to Fezf2 in Vitro—To purify DNA fragments bound by Fezf2, we generated a recombinant Fezf2 ZF domain fused to GST, which was immobilized to a solid surface and used to screen total zebrafish genomic DNA digested with the restriction endonuclease Sau3A1. The fragments bound by Fezf2 were subsequently eluted and amplified by PCR. Four rounds of selection and PCR amplification were performed (Fig. 1). PCR products from the second, third, and fourth rounds were cloned and sequenced.

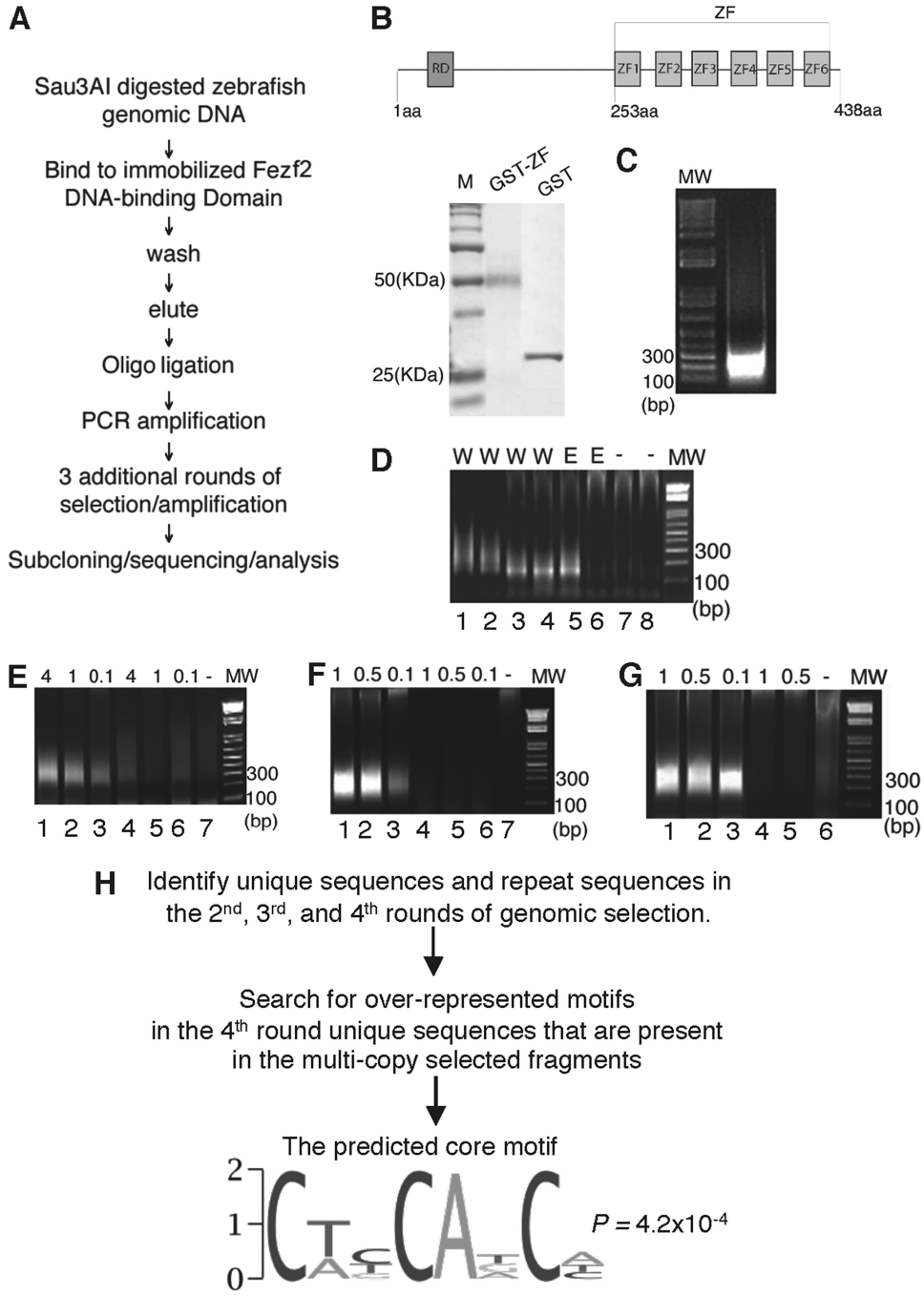

FIGURE 1. ***In vitro* genomic selection and computational prediction uncovers a putative Fezf2 core binding site.** *A*, flow chart of the *in vitro* genomic selection method. *B*, schematic diagram of the Fezf2 protein and Coomassie Blue-stained gel image showing the expression of GST-Fezf2 zinc finger domain fusion (GST-ZF) as well as the GST control. *C*, ethidium bromide-stained gel image showing the Sau3A1-digested zebrafish genomic DNA. *D–G*, PCR amplification of the genomic fragments after each round of selection. *D*, first PCR. Prior to the first PCR, 1 pmol of NotI/Sau3A oligonucleotides were ligated to either elution (*E*, *lanes 5* and *6*) or 0.5 M salt wash (*W*, *lanes 1-4*) fractions from the first selection round. *Lanes 1*, *3*, and *5* are samples from the column coupled with the GST-ZF. *Lanes 2*, *4*, and *6* are samples from the column coupled with GST control. 1–2 ml of the PCR product (as shown in *lane 5 or 6*) was used for subsequent rounds of selection. *Lanes 7* and *8* are PCR negative control (*lane 7*, PCR with only NotI/Sau3A1 oligonucleotides as templates; *lane 8*, PCR with only primers). *E–G*, Second to fourth rounds of PCR. *Lanes 1–3* are samples selected with GST-ZF, and *lanes 4-6* are those selected with GST control. *Lane 7* (*E* and *F*) and *lane 6* (*G*) are PCR negative controls. The amount of *in vitro* selected genomic templates used for PCR is indicated above the gel images. *H*, schematic diagram showing the procedure for computational prediction of the core motif.

BLAST search revealed that, among the ~800 sequenced clones, 220 aligned with specific sequences in the zebrafish genome thus representing unique genomic fragments, whereas the rest had multiple (greater than 5) hits, suggesting that they likely correspond to repetitive sequences (Table 1).

Computational Analysis Identifies a Core Consensus Fezf2 Binding Site—Among the 220 unique genomic fragments, 6 were represented multiple times, especially among those from the fourth round of selection, whereas others were singletons. To uncover the core consensus Fezf2 binding site(s), we rea-

TABLE 1
Number of sequences in each round and percentage of unique fragments

Sequences	Second round	Third round	Fourth round
Number of total sequences	190	194	411
Number of unique sequences	47	56	117
Percentage of unique sequences	24.7%	28.8%	28.5%

soned that the fragments truly bound by Fezf2 would become enriched through each round of selection, regardless of whether they are present as singletons or not. Several motif-finding computational algorithms such as BioProspector, AlignACE, and MEME (21–23) were therefore used to identify sequence motifs that are enriched in the unique fragments from the fourth round. Because the unique fragments that were identified multiple times are most likely bound by Fezf2, we considered only the predicted motifs that appeared in this group. The top motifs of different widths that were predicted using BioProspector are similar (supplemental Table S1). Moreover, the top motifs identified by AlignACE and MEME were also similar to those identified with BioProspector (supplemental Tables S2 and S3). Together, these motif-finding analyses employing different algorithms suggest the CnnCAnCn core as a putative consensus Fezf2 binding site (Fig. 1*H*).

Core Consensus Fezf2 Binding Site Unveils Response Elements That Interact with Fezf2 in Vitro—We searched the selected genomic fragments for the computationally predicted Fezf2 core binding site and determined the ability of Fezf2 to interact with these putative response elements *in vitro*. We first used fluorescence anisotropy, which is a rapid, sensitive, and quantitative technique for the analysis of protein-protein and protein-DNA interactions in solution, through measuring the depolarization of emitted fluorescence intensity obtained after excitation by a polarized light source. Binding of an unlabeled macromolecule (*e.g.* protein) can change the tumbling time of DNA to which the fluorescent probe is attached (*e.g.* DNA oligonucleotides) and hence the measured anisotropy (24, 25). Using this method, we found that Fezf2 bound with high affinity (in the nanomolar range) to oligonucleotides bearing the putative Fezf2 response elements (Fig. 2, *A* and *B*). The K_d of Fezf2 binding to the response element 4.26 (containing two Fezf2 binding sites) is less than half of that of 4.430 (containing one Fezf2 binding site), suggesting a possible cooperative binding of Fezf2 to 4.26. Mutating the consensus nucleotides in the Fezf2 response elements significantly impaired the binding (Fig. 2*B*).

We next asked whether Fezf2 regulates the transcription of luciferase reporter constructs linked to multiple copies of Fezf2 response elements (Fig. 2*C*) in cultured mammalian cells. Zebrafish Fezf2 in HEK293 cells enhanced the activation of luciferase by 1.5–4-fold, whereas mutating the consensus residues in the Fezf2 response elements significantly compromised this transactivation (Fig. 2*D*). We also tested whether the mouse Fezf2 could transactivate the luciferase reporter through the same response elements. Mouse Fezf2 transactivated the luciferase reporter in three of four cases tested, and such activity was significantly impaired when the consensus residues in the Fezf2 response elements were mutated (Fig. 2*E*). Together, these results show that *in vitro* genomic selection followed by computational prediction can uncover sequence elements that interact directly with both zebrafish and mouse Fezf2 *in vitro*.

Fezf2 Response Elements Interact with Fezf2 in Vivo—To determine whether Fezf2 interacts with its response elements *in vivo*, we cloned individual response elements into the Tol2 transposon vector, proximal to the minimal promoter E1b and the EGFP reporter, and injected the constructs into zebrafish embryos (Fig. 3*A*). With all four response elements tested, we observed GFP signals that were largely restricted to the forebrain (Fig. 3*B*; statistics indicated in the figure panels). In agreement with previous observations in zebrafish (26) or in mice (27), although these kinds of short sequences do display enhancer activity, they tend to yield results with more mosaicism and less penetrance. Nevertheless, these results suggest that a single Fezf2 response element is sufficient to act as a forebrain enhancer in zebrafish.

To verify whether this *in vivo* forebrain enhancer activity is truly dependent on the response elements, we changed the absolutely conserved three C nucleotides in the wild-type motif CAGCAACC, thereby generating an altered element TAGAAAAC. This mutated element had no forebrain enhancer activity (Fig. 3*C*, compare the *middle column* with the *left column*; statistics indicated in the figure panels). We further determined whether the forebrain enhancer activity of the wild-type motif CAGCAACC is dependent on Fezf2. By injecting the wild-type motif-driven reporter construct into the *fezf2* morphant, which has significantly impaired Fezf2 activity (14, 28), we detected no GFP signal (Fig. 3*C*, compare the *right column* with the *left column*; statistics indicated in the figure panels), indicating that *fezf2* gene activity is essential for the motif to exert a forebrain enhancer activity. Together, these results indicate that the core consensus binding site identifies Fezf2 response elements that interact with Fezf2 *in vivo*.

Analyses of Selected Unique Genomic Fragments Bearing Fezf2 Core Consensus Binding Sites Reveal High Degree of Conservation and an Unusual Overlap with Protein-coding Exons—Having established that the computationally predicted Fezf2 core consensus binding site identifies *bona fide* Fezf2 response elements both *in vitro* and *in vivo*, we carried out further analyses of the selected unique genomic DNA fragments bound by Fezf2. An analysis of the Fezf2 binding site distribution in the genome uncovered a frequency of 0.392% (3.92 motifs/1000 bp) in repetitive sequences and 0.423% in nonrepetitive sequence. The frequency of the Fezf2 binding site in the SELEXed unique and repetitive fragments is 0.8% and 0.707%, respectively, both of which represent significant enrichment over the genome-wide distribution. We found that ~75% of the fourth round fragments contained at least one core binding site, compared with ~66 and ~51% of the second and third rounds, respectively (Fig. 4*A*). Among the fragments containing the core binding site, a majority of them had one or two core binding sites, whereas a small percentage had more than four binding sites (Fig. 4*B*). Additionally, ~50% of the fragments represented evolutionarily conserved sequences according to the UCSC genome browser (Fig. 4*C*). This represents a significant enrichment (the p value is $8.6e^{-5}$) over the background probability (which is calculated to be ~0.35) of having a conserved fragment by random sampling.

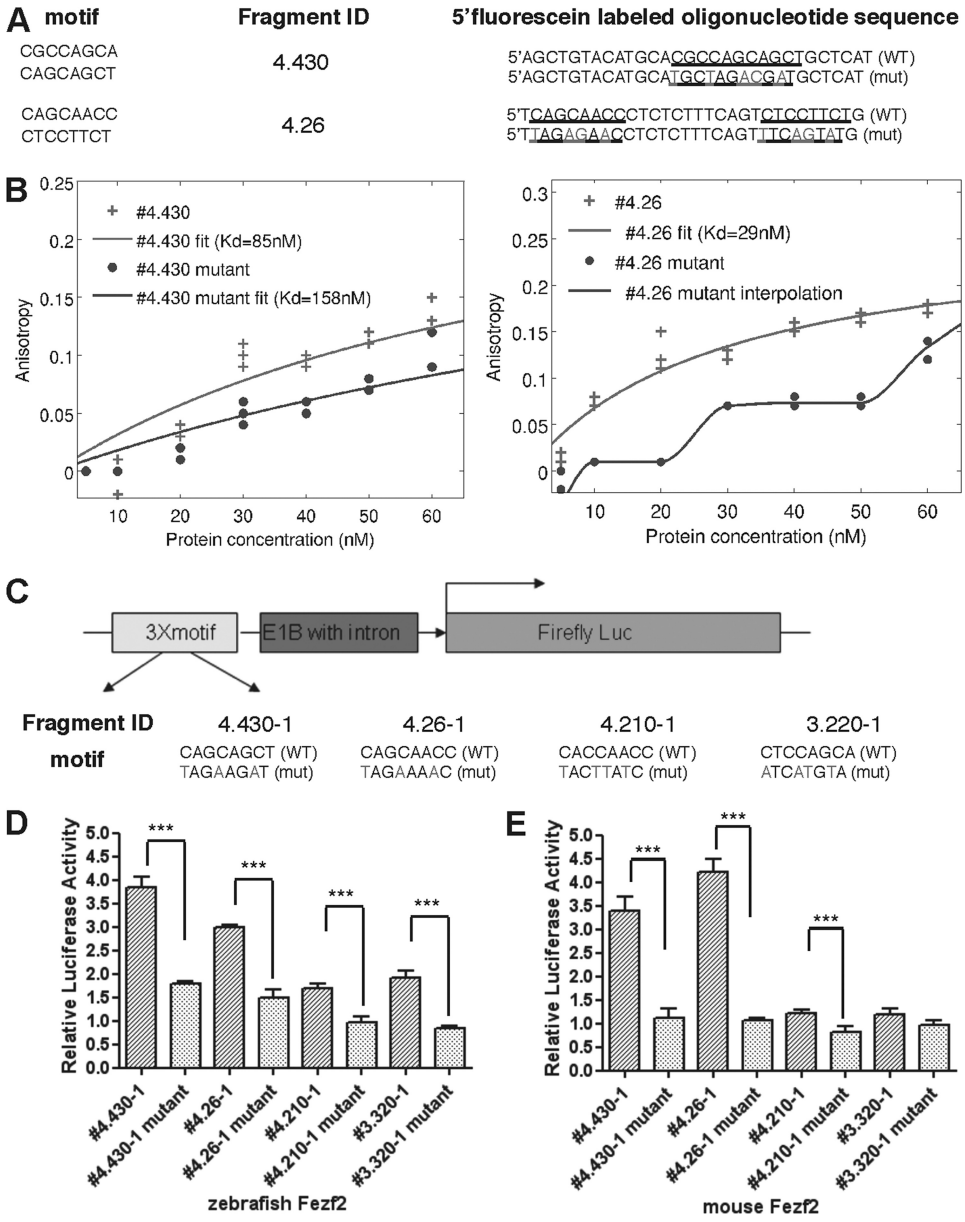

FIGURE 2. **Computationally predicted core motif functions as Fezf2 response elements *in vitro*.** *A*, fluorescein-labeled double-stranded oligonucleotides carrying the WT or mutated predicted response elements (0.1 nM) from *in vitro* selected genomic fragments 4.26 and 4.430 are shown. Motifs contained in these sequences are *underlined* as well as shown on the *left*. *B*, binding of GST-ZF to these elements was measured by fluorescent anisotropy, and the anisotropy graphs are shown. K_dd values were calculated from curve fits. For the 4.26 mutant sequence, the *curve* is drawn by interpolation. *C*, schematic diagram shows the constructs used for the luciferase assay. *D* and *E*, Dual-Luciferase assay of zebrafish Fezf2 (*D*) and mouse Fezf2 (*E*) show that both Fezf2 can activate the luciferase reporter driven by WT response elements. Mutating the conserved residues in the response elements significantly impairs the transactivation. Relative luciferase activity is calculated as a ratio of Fezf transfected to control plasmid transfected, after normalization of firefly luciferase to *Renilla* luciferase activity. ***, $p < 0.001$. *Error bars*, S.E.

We also mapped these genomic fragments to the zebrafish genome. Surprisingly, we found that ~20–30% of the genomic fragments overlapped well annotated protein-coding exons (Fig. 4*D*). Encompassing this group, about half of the fragments resided in protein-coding genes (*i.e.* they overlapped the gene start/stop position information from the Ensembl database), whereas a small percentage of fragments resided >50 kb away from protein-coding genes (Fig. 4*E*). Thus, our analyses reveal a high degree of conservation and an unusual overlap with protein-coding exons in selected unique genomic fragments bearing Fezf2 core consensus binding site(s).

Evolutionarily Conserved Transcription Regulators Eomesa/Tbr2 and Lhx2 Are Direct Targets of Fezf2 in Forebrain Development—To determine whether the selected genomic fragments bound by Fezf2 would lead us to the identification of *bona fide* Fezf2 target genes, we carried out in-depth analysis of two genomic fragments, 4.430 and 4.210, because of their close proximity to known forebrain-expressed genes. The fragment 4.430 resides ~16 kb upstream of the *eomesa/tbr2* gene (Fig. 5*A*), which is expressed in the developing forebrain and critical for regulating forebrain patterning and neurogenesis (29–31). The fragment 4.210 resides ~8 kb upstream of the *lhx2b* gene

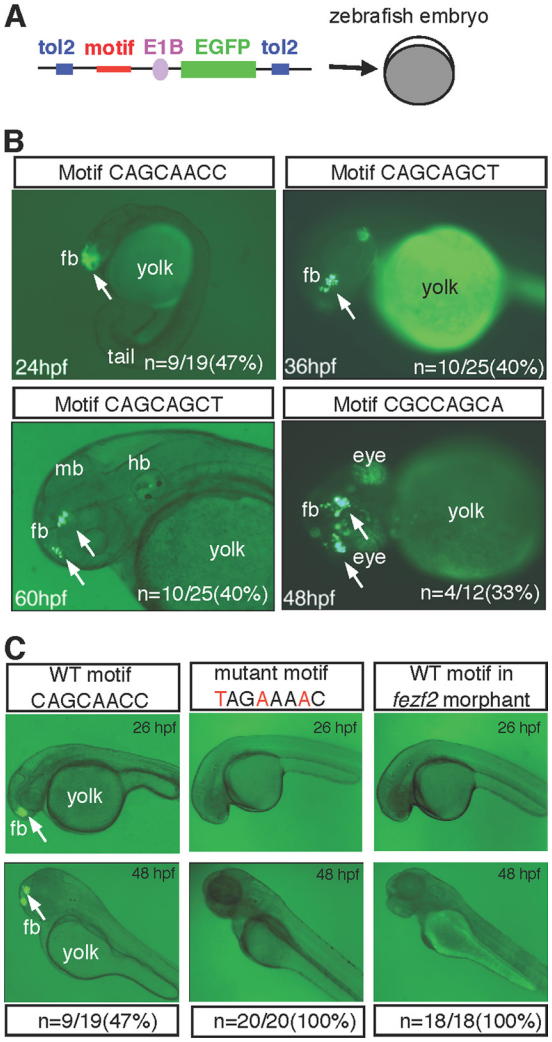

FIGURE 3. **Computationally predicted core motif functions as Fezf2-response element *in vivo*.** *A*, schematic diagram shows the *in vivo* reporter assay construct. *B*, images of transient transgenic zebrafish embryos show that four response elements residing in *in vitro*-selected genomic fragments can serve as forebrain enhancers *in vivo*. All are lateral views except the *last panel* (ventral view). *C*, images of transient transgenic zebrafish embryos show that mutating the conserved residues in the core binding site abolishes forebrain enhancer activity (*middle column*), and impairing Fezf2 activity through morpholino antisense oligonucleotide-mediated knockdown also impairs forebrain enhancer activity (*right column*).

(Fig. 5*B*), which is also expressed in the developing forebrain and plays an important role in forebrain patterning, cell fate specification, and axonal guidance (32–34).

To determine whether Fezf2 interacts with these genomic fragments *in vivo*, we performed *in vivo* ChIP, using a custom-generated polyclonal antibody that specifically detected zebrafish Fezf2 protein in Western blot analysis (Fig. 5*C*). This Fezf2 antibody was able to pull down genomic fragments near *eomesa/tbr2* and *lhx2b*, whereas it did not pull down an unrelated genomic fragment (Fig. 5*D*). Also, control IgG did not pull down any genomic fragments as shown (Fig. 5*D*). These results suggest that Fezf2 interacts with genomic fragments in the vicinity of *eomesa/tbr2* and *lhx2b* genes *in vivo*.

In zebrafish, the expression of *eomesa* was detected beginning at ~10-somite stage in the telencephalon (Fig. 5*E*, *middle column*). At ~24 and 36 h post-fertilization (hpf), *eomesa* expression was detected in both the posterior telencephalon and prethalamus, where *fezf2* expression was also detected (Fig. 5*E*, *middle column*). *lhx2b* expression was detected broadly in many domains of the nervous system at ~10-somite stage. At ~24 hpf, *lhx2b* expression was highly enriched in the telencephalon as well as the anterior diencephalon, where it overlapped with *fezf2* expression. At ~36 hpf, *lhx2b* expression became down-regulated but remained detectable in the telencephalon and anterior diencephalon in overlapping regions with that of *fezf2* (Fig. 5*E*, *right column*). To determine whether Fezf2 indeed regulates *eomesa* and *lhx2b in vivo*, we examined their expression in the *fezf2* mutant, morphant, and overexpression (via *hsp-gal4;uas-fezf2*) conditions. In the *too few* (*tof*) mutant, which changes a single conserved cysteine to serine in the ZF domain (13), *eomesa* expression in both the telencephalon and diencephalon was reduced (Fig. 5*F*, *top row*, *second panel*, $n > 8$). Knockdown of *fezf2* activity with a specific morpholino antisense oligonucleotide yielded similar results (Fig. 5*F*, *top row*, *third panel*, $n > 20$). These data indicate that *fezf2* is necessary to regulate *eomesa* expression. Conversely, overexpression of *fezf2* via the use of *hsp-gal4;uas-fezf2* double transgenic animals (14) led to a significant expansion of *eomesa*-expressing domain in both the telencephalon and diencephalon (Fig. 5*F*, *top row*, *fourth panel*, $n > 8$). Together, these analyses indicate that *fezf2* is necessary and sufficient in certain cellular contexts to regulate *eomesa* expression *in vivo*.

In the *tof* mutant and *fezf2* morphant, *lhx2b* expression was selectively reduced in the forebrain but not in the mid-hindbrain regions (Fig. 5*F*, *bottom row*, *second* and *third panels*, $n > 8$ and $n > 20$, respectively), consistent with a dependence on *fezf2* to activate or maintain its expression in the forebrain. Overexpression of *fezf2* via the *hsp-gal4;uas-fezf2* system led to significant up-regulation of *lhx2b* expression throughout the embryo (Fig. 5*F*, *bottom row*, *fourth panel*). Thus, *fezf2* is necessary and sufficient in certain cellular contexts to regulate *lhx2b* expression *in vivo*.

Loss of *lhx2b* function in the zebrafish mutant *belladonna* causes a midline axon guidance defect (34). Such axonal guidance defect has not been previously reported in the *tof* mutant. Because *fezf2* is critical to regulating *lhx2b* expression, we wondered whether the *tof* mutant might also suffer an axonal guidance defect. Immunostaining with the anti-acetylated tubulin antibody uncovered that the anterior commissure as well as the post-optic commissure failed to cross the midline in both the *tof* mutant and *fezf2* morphant properly (Fig. 5*G*) at 36 hpf. However, this defect did recover later, such that no obvious axonal targeting defect was observed in the 48 hpf *tof* mutant and *fezf2* morphant. Together, these results suggest that Fezf2 is required for proper midline axonal guidance through activating or maintaining the expression of *lhx2b*.

DISCUSSION

In this study, we extend *in vitro* genomic SELEX, previously applicable only in simple organisms with smaller genomes (8, 9), to a complex vertebrate genome. This method is potentially applicable to any DNA-binding protein for which a functional DNA binding domain can be purified. Compared with SELEX using short random oligonucleotides, genomic SELEX leads to

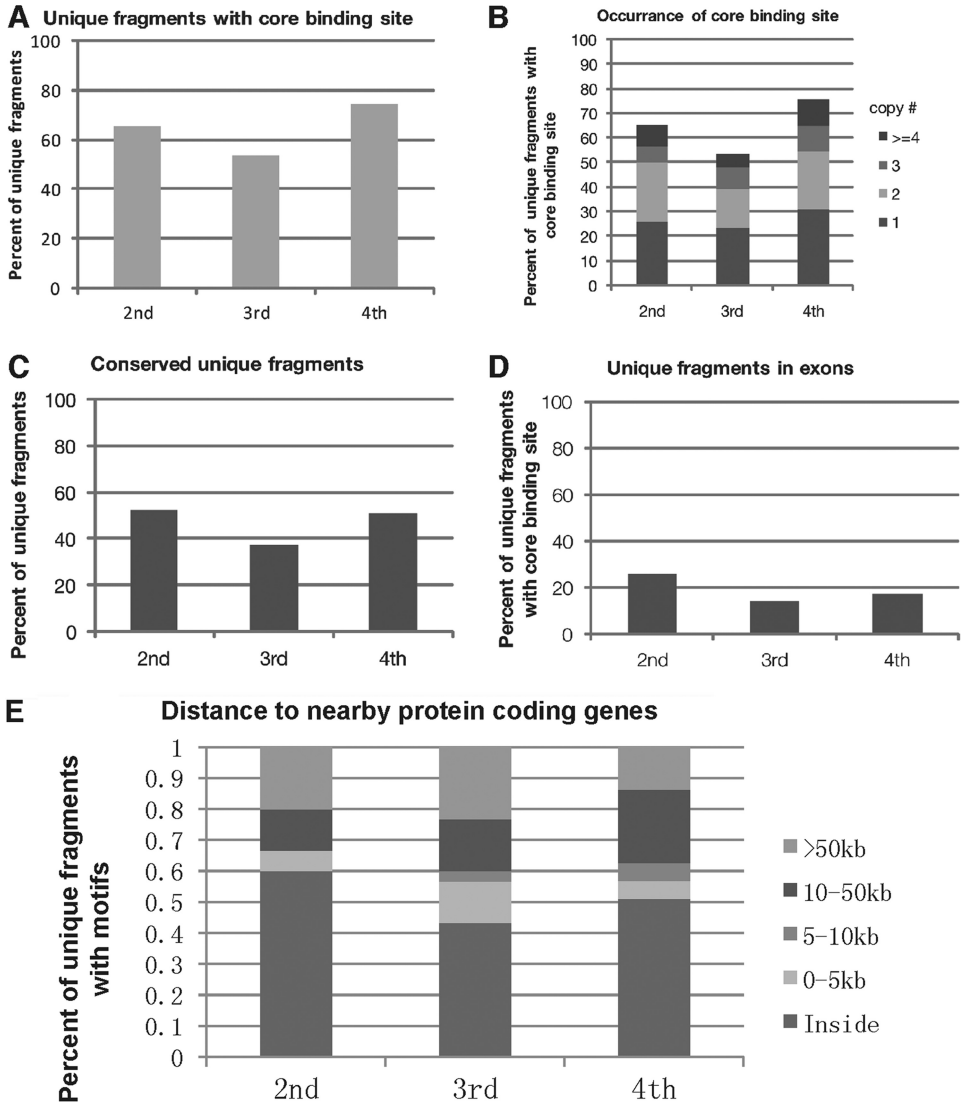

FIGURE 4. **Analyses of the selected unique genomic fragments.** *A*, percent of unique genomic fragment containing at least one core binding site. *B*, occurrence of core binding sites in 1, 2, 3, and 4 or greater copies. *C*, percent of unique fragments that are evolutionarily conserved between zebrafish and human. *D*, percent of unique genomic fragments that are annotated as protein-coding exons. *E*, diagram showing the frequency among the selected unique genomic fragments containing the core binding sites, which are located either inside the gene, or 0–5 kb, 5–10 kb, 10–50 kb, or >50 kb from a gene.

identification of genomic sites of interaction, which provides a basis for identifying direct target genes and building gene networks. Furthermore, we define the DNA binding sites for the evolutionarily conserved forebrain regulatory protein Fezf2 and show that the binding sites function as Fezf2-response elements both *in vitro* and *in vivo*. Finally, analyses of two target genes reveal a connection to human congenital microcephaly syndromes and a new role of Fezf2 in guiding zebrafish commissural axons.

Because the Fezf2 binding site reported in this study was recovered by searching for enriched motifs in the fourth round fragments, this binding site may represent a high affinity binding site for Fezf2. To look for a potentially existing low affinity Fezf2 binding site, one can possibly search for motifs enriched in the second round but not in the fourth round. Thus, *in vitro* genomic selection is potentially applicable for identifying both low and high affinity binding sites.

Approximately one-quarter of the selected genomic fragments bound by Fezf2 represent unique sites in the genome. Although this percentage is lower than that obtained in *C. elegans* (9), it is not unexpected, because the vertebrate genome is known to contain a much larger amount of repetitive elements than the invertebrate genome. Indeed, an analysis of the zebrafish genome shows that ~53% of the sequences are marked by "repeat mask" as repetitive sequences. Given that the SELEX method involves multiple rounds of PCR amplification, it is not surprising that many of the selected fragments are from repetitive sequences. Because both target genes were validated in this study, we suggest that the presence of repetitive sequences, although unavoidable, does not impact the physiological relevance of the selected unique sequences. A majority of the selected unique genomic fragments remain singletons after four rounds of selection and amplification. This could be due to the size and complexity of the vertebrate genome; alternatively, it could suggest the presence of many direct target genes of Fezf2.

It is interesting to note that ~20% of the selected unique genomic fragments bound by Fezf2 overlap well annotated pro-

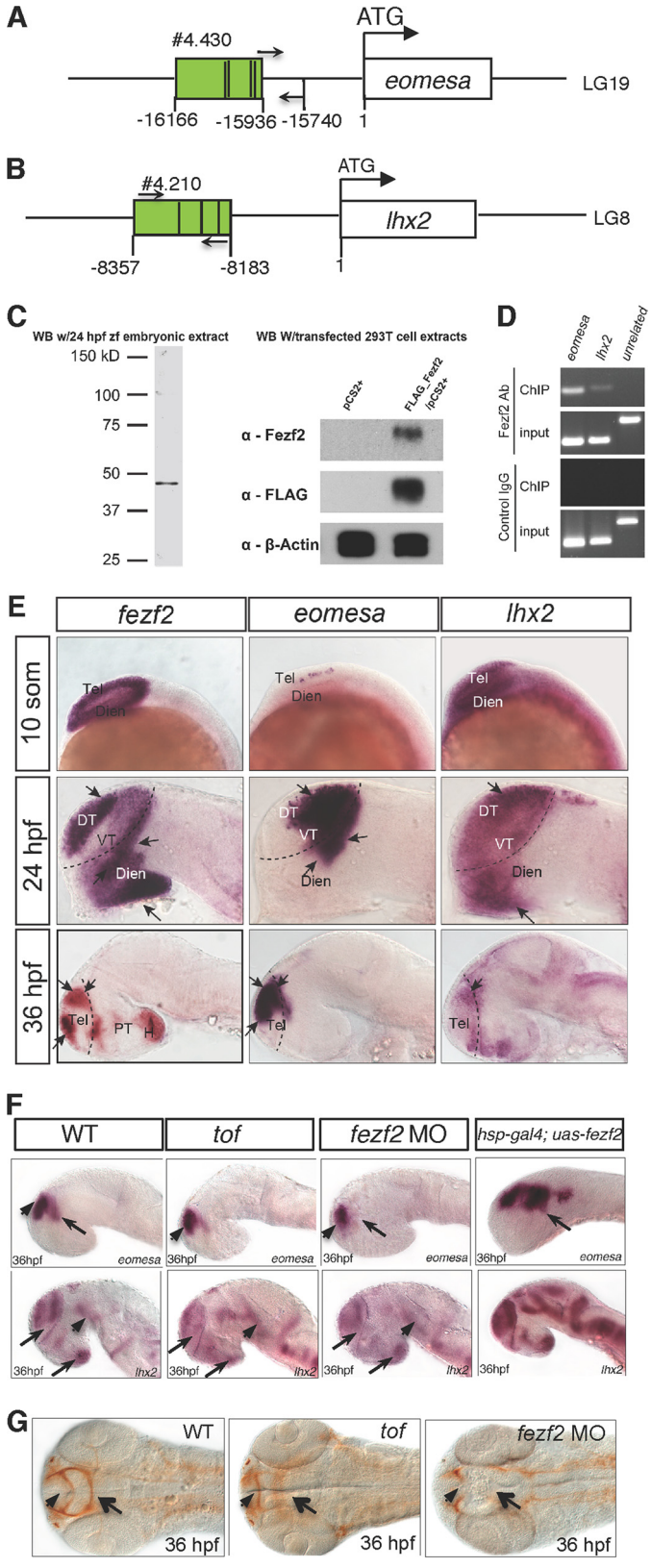

FIGURE 5. **Fezf2 directly activates *eomesa* and *lhx2b in vivo*.** *A* and *B*, schematic diagrams show the location of the Fezf2 binding sites (*black bars*) in selected genomic fragments (*green*) located upstream of *eomesa* (*A*) and *lhx2b* (*B*) genes. *C*, Western blotting shows that the affinity-purified Fezf2 antibody specifically recognizes the Fezf2 protein in zebrafish embryonic extracts and in transfected HEK293 cells. *D*, ChIP analysis shows that Fezf2 binds to the response elements near *eomesa* and *lhx2b in vivo*. ChIP primers are indicated in *A* and *B* (*arrows*). The unrelated genomic fragment is a computationally identified zebrafish genomic fragment that contains no Fezf2 binding site. *E*, *in situ* hybridization shows the expression of *fezf2*, *eomesa*, *and lhx2b* at three developmental stages. *Fezf2* expression in the forebrain precedes and overlaps with that of *eomesa* and *lhx2b*. *F*, *in situ* hybridization shows the expression of *eomesa* (*first row*) and *lhx2b* (*second row*) in WT (*left column*), the *tof* mutant (*second column*), the *fezf2* morphant (*third column*), and the *fezf2*-overexpressing (*hsp-gal4;uas-fezf2*) embryos (*right column*). Fezf2 is necessary and sufficient to regulate the expression of *eomesa* and *lhx2b*. *G*, labeling with the anti-acetylated tubulin antibody shows the defect of commissural axon targeting in the *tof* mutant and *fezf2* morphant.

tein-coding exons. Although protein coding and regulatory functions are conventionally treated as nonoverlapping, some recent studies begin to challenge this view and suggest that exons can also carry regulatory function (35–37). It will be of great interest in the future to test for potential enhancer activity of these exonic sequences that are bound by Fezf2.

Our in-depth analyses of two selected genomic fragments reveal that Fezf2 directly activates evolutionarily conserved genes *eomesa/tbr2* and *lhx2b*. The reduction rather than complete abolishment of their expression in *fezf2*-deficient conditions may be due to partial inactivation of Fezf2 in the mutant and morphant, or it may reflect partial dependence of their expression on Fezf2. Eomesa/Tbr-2 is a T-box transcription regulator, human mutations of which lead to an autosomal recessive microcephaly syndrome (29). In the mouse brain, Eomesa/Tbr2 is required for the specification of cortical intermediate progenitor cells (30, 31). Thus, the role of Fezf2 in forebrain patterning and cell fate specification is likely mediated in part through direct activation of *eomesa* transcription. Given the conservation of Fezf2 and its binding sites across vertebrates, it is plausible that Fezf2 may directly regulate *eomesa* in mammals, making *fezf2* a good candidate gene to be evaluated in disorders of microcephaly. Our findings also reveal that Fezf2 directly regulates *lhx2b*, which is required for mouse forebrain patterning and eye formation (32) and was recently shown to have a selector activity in specifying cortical cell identity (33). In zebrafish, *lhx2b/belladonna* is required for neural patterning and midline axon guidance in the forebrain (34). Our analysis reveals a previously unknown role of Fezf2 in outgrowth and targeting of commissural axons in zebrafish, likely mediated at least in part by its regulation of *lhx2b*.

It is possible that there are many more direct target genes of Fezf2 than we currently know, as for example, a recent study in mice shows that Fezf2 can directly repress Hes5 (38). Moreover, the target genes of Fezf2 in the development of dopaminergic and serotonergic neurons remain to be identified. It would therefore be of interest to analyze additional predicted target genes as well as to further refine the DNA binding property of Fezf2, for example, by characterizing the nature of flanking sequences using recently developed methods (39). Finally, it is of interest to discuss whether genomic SELEX provides additional advantage over sequence-based data mining even under the condition that the TF binding sites are already known. Future work to validate all of the identified 220 Fezf2-bound unique genomic fragments as well as to identify more selected fragments by deep sequencing and compare them with sequence-based data mining will allow this question to be addressed.

Acknowledgments—We thank M. Hibi, M. Jamrich, and R. Karlstrom for plasmids; the K. Yamamoto laboratory members for advice on in vitro genomic selection and fluorescence anistropy; the D. Sheppard laboratory members for the use of Tecan fluorometer; the Guo laboratory members for helpful discussions; and Z. Dong, M. Berberoglu, and B. Lu for reading the manuscript.

REFERENCES

1. Iyer, V. R., Horak, C. E., Scafe, C. S., Botstein, D., Snyder, M., and Brown, P. O. (2001) *Nature* **409,** 533–538
2. Johnson, D. S., Mortazavi, A., Myers, R. M., and Wold, B. (2007) *Science* **316,** 1497–1502
3. Blackwell, T. K., and Weintraub, H. (1990) *Science* **250,** 1104–1110
4. Ellington, A. D., and Szostak, J. W. (1990) *Nature* **346,** 818–822
5. Tuerk, C., and Gold, L. (1990) *Science* **249,** 505–510
6. Djordjevic, M. (2007) *Biomol. Eng.* **24,** 179–189
7. Jolma, A., Kivioja, T., Toivonen, J., Cheng, L., Wei, G., Enge, M., Taipale, M., Vaquerizas, J. M., Yan, J., Sillanpää, M. J., Bonke, M., Palin, K., Talukder, S., Hughes, T. R., Luscombe, N. M., Ukkonen, E., and Taipale, J. (2010) *Genome Res.* **20,** 861–873
8. Shtatland, T., Gill, S. C., Javornik, B. E., Johansson, H. E., Singer, B. S., Uhlenbeck, O. C., Zichi, D. A., and Gold, L. (2000) *Nucleic Acids Res.* **28,** E93
9. Shostak, Y., Van Gilst, M. R., Antebi, A., and Yamamoto, K. R. (2004) *Genes Dev.* **18,** 2529–2544
10. Matsuo-Takasaki, M., Lim, J. H., Beanan, M. J., Sato, S. M., and Sargent, T. D. (2000) *Mech. Dev.* **93,** 201–204
11. Hashimoto, H., Yabe, T., Hirata, T., Shimizu, T., Bae, Y., Yamanaka, Y., Hirano, T., and Hibi, M. (2000) *Mech. Dev.* **97,** 191–195
12. Guo, S., Wilson, S. W., Cooke, S., Chitnis, A. B., Driever, W., and Rosenthal, A. (1999) *Dev. Biol.* **208,** 473–487
13. Levkowitz, G., Zeller, J., Sirotkin, H. I., French, D., Schilbach, S., Hashimoto, H., Hibi, M., Talbot, W. S., and Rosenthal, A. (2003) *Nat. Neurosci.* **6,** 28–33
14. Jeong, J. Y., Einhorn, Z., Mathur, P., Chen, L., Lee, S., Kawakami, K., and Guo, S. (2007) *Development* **134,** 127–136
15. Berberoglu, M. A., Dong, Z., Mueller, T., and Guo, S. (2009) *Gene Expr. Patterns* **9,** 411–422
16. Rouaux, C., and Arlotta, P. (2010) *Nat. Neurosci.* **13,** 1345–1347
17. Hirata, T., Suda, Y., Nakao, K., Narimatsu, M., Hirano, T., and Hibi, M. (2004) *Dev. Dyn.* **230,** 546–556
18. Molyneaux, B. J., Arlotta, P., Hirata, T., Hibi, M., and Macklis, J. D. (2005) *Neuron* **47,** 817–831
19. Chen, B., Schaevitz, L. R., and McConnell, S. K. (2005) *Proc. Natl. Acad. Sci. U.S.A.* **102,** 17184–17189
20. Shimizu, T., and Hibi, M. (2009) *Dev. Growth Differ.* **51,** 221–231
21. Liu, X., Brutlag, D. L., and Liu, J. S. (2001) *Pac. Symp. Biocomput.* 127–138
22. Roth, F. P., Hughes, J. D., Estep, P. W., and Church, G. M. (1998) *Nat. Biotechnol.* **16,** 939–945
23. Bailey, T. L., Williams, N., Misleh, C., and Li, W. W. (2006) *Nucleic Acids Res.* **34,** W369–373
24. Chin, J., Längst, G., Becker, P. B., and Widom, J. (2004) *Methods Enzymol.* **376,** 3–16
25. Anderson, B. J., Larkin, C., Guja, K., and Schildbach, J. F. (2008) *Methods Enzymol.* **450,** 253–272
26. Li, Q., Ritter, D., Yang, N., Dong, Z., Li, H., Chuang, J. H., and Guo, S. (2010) *Dev. Biol.* **337,** 484–495
27. De Val, S., Chi, N. C., Meadows, S. M., Minovitsky, S., Anderson, J. P., Harris, I. S., Ehlers, M. L., Agarwal, P., Visel, A., Xu, S. M., Pennacchio, L. A., Dubchak, I., Krieg, P. A., Stainier, D. Y., and Black, B. L. (2008) *Cell* **135,** 1053–1064
28. Jeong, J. Y., Einhorn, Z., Mercurio, S., Lee, S., Lau, B., Mione, M., Wilson, S. W., and Guo, S. (2006) *Proc. Natl. Acad. Sci. U.S.A.* **103,** 5143–5148
29. Baala, L., Briault, S., Etchevers, H. C., Laumonnier, F., Natiq, A., Amiel, J., Boddaert, N., Picard, C., Sbiti, A., Asermouh, A., Attié-Bitach, T., Encha-Razavi, F., Munnich, A., Sefiani, A., and Lyonnet, S. (2007) *Nat. Genet.* **39,** 454–456
30. Sessa, A., Mao, C. A., Hadjantonakis, A. K., Klein, W. H., and Broccoli, V. (2008) *Neuron* **60,** 56–69
31. Arnold, S. J., Huang, G. J., Cheung, A. F., Era, T., Nishikawa, S., Bikoff, E. K., Molnár, Z., Robertson, E. J., and Groszer, M. (2008) *Genes Dev.* **22,** 2479–2484
32. Porter, F. D., Drago, J., Xu, Y., Cheema, S. S., Wassif, C., Huang, S. P., Lee, E., Grinberg, A., Massalas, J. S., Bodine, D., Alt, F., and Westphal, H. (1997) *Development* **124,** 2935–2944
33. Mangale, V. S., Hirokawa, K. E., Satyaki, P. R., Gokulchandran, N., Chikbire, S., Subramanian, L., Shetty, A. S., Martynoga, B., Paul, J., Mai, M. V., Li, Y., Flanagan, L. A., Tole, S., and Monuki, E. S. (2008) *Science* **319,** 304–309
34. Seth, A., Culverwell, J., Walkowicz, M., Toro, S., Rick, J. M., Neuhauss, S. C., Varga, Z. M., and Karlstrom, R. O. (2006) *Development* **133,** 725–735
35. Lampe, X., Samad, O. A., Guiguen, A., Matis, C., Remacle, S., Picard, J. J., Rijli, F. M., and Rezsohazy, R. (2008) *Nucleic Acids Res.* **36,** 3214–3225
36. Tümpel, S., Cambronero, F., Sims, C., Krumlauf, R., and Wiedemann, L. M. (2008) *Proc. Natl Acad. Sci. U.S.A.* **105,** 20077–20082
37. Dong, X., Navratilova, P., Fredman, D., Drivenes, Ø., Becker, T. S., and Lenhard, B. (2010) *Nucleic Acids Res.* **38,** 1071–1085
38. Shimizu, T., Nakazawa, M., Kani, S., Bae, Y. K., Shimizu, T., Kageyama, R., and Hibi, M. (2010) *Development* **137,** 1875–1885
39. Fordyce, P. M., Gerber, D., Tran, D., Zheng, J., Li, H., DeRisi, J. L., and Quake, S. R. (2010) *Nat. Biotech.* **28,** 970–975
40. Westerfield, M. (ed) (1995) *The Zebrafish Book: A Guide for the Laboratory Use of Zebrafish, Brachydanio rerio*, 3rd Ed., University of Oregon Press, Eugene, OR
41. Fujita, P. A., Rhead, B., Zweig, A. S., Hinrichs, A. S., Karolchik, D., Cline, M. S., Goldman, M., Barber, G. P., Clawson, H., Coelho, A., Diekhans, M., Dreszer, T. R., Giardine, B. M., Harte, R. A., Hillman-Jackson, J., Hsu, F., Kirkup, V., Kuhn, R. M., Learned, K., Li, C. H., Meyer, L. R., Pohl, A., Raney, B. J., Rosenbloom, K. R., Smith, K. E., Haussler, D., and Kent, W. J. (2011) *Nucleic Acids Res.* **39,** D876–882
42. Siepel, A., Bejerano, G., Pedersen, J. S., Hinrichs, A. S., Hou, M., Rosenbloom, K., Clawson, H., Spieth, J., Hillier, L. W., Richards, S., Weinstock, G. M., Wilson, R. K., Gibbs, R. A., Kent, W. J., Miller, W., and Haussler, D. (2005) *Genome Res.* **15,** 1034–1050
43. Krishnan, K., Salomonis, N., and Guo, S. (2008) *PLoS ONE* **3,** e3621
44. Chen, H., Contreras, X., Yamaguchi, Y., Handa, H., Peterlin, B. M., and Guo, S. (2009) *PLoS One* **4,** e6918
45. Guo, S., Brush, J., Teraoka, H., Goddard, A., Wilson, S. W., Mullins, M. C., and Rosenthal, A. (1999) *Neuron* **24,** 555–566