Identification of Distal cis-Regulatory Elements at Mouse Mitoferrin Loci Using Zebrafish Transgenesis

The Harvard community has made this article openly available. Please share how this access benefits you. Your story matters.

Citation

Published Version
doi:10.1128/MCB.01010-10

Citable link
http://nrs.harvard.edu/urn-3:HUL.InstRepos:17085600

Terms of Use
This article was downloaded from Harvard University’s DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA
Identification of Distal cis-Regulatory Elements at Mouse Mitoferrin Loci Using Zebrafish Transgenesis

Julio D. Amigo,1|| Ming Yu,2† Marie-Berengere Troadeç,3‡ Babette Gwynn,4 Jeffrey D. Cooney,1 Amy J. Lambert,1§ Neil C. Chi,5 Mitchell J. Weiss,6 Luanne L. Peters,4§ Jerry Kaplan,5 Alan B. Cantor,2 and Barry H. Paw1,2*

Department of Medicine, Division of Hematology, Brigham and Women's Hospital, and Harvard Medical School, Boston, Massachusetts 021151; Department of Medicine, Division of Hematology-Oncology, Children's Hospital, Boston, and Harvard Medical School, Boston, Massachusetts 021152; Department of Pathology, University of Utah School of Medicine, Salt Lake City, Utah 841323; Jackson Laboratory, Bar Harbor, Maine 046094; Department of Medicine, Division of Cardiology, University of California, San Diego, School of Medicine, La Jolla, California 92037; and Division of Hematology, Children's Hospital of Philadelphia, and University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania 191046

Received 29 August 2010/Returned for modification 19 October 2010/Accepted 6 January 2011

Mitoferrin 1 (Mfrn1; Slc25a37) and mitoferrin 2 (Mfrn2; Slc25a28) function as essential mitochondrial iron importers for heme and Fe/S cluster biogenesis. A genetic deficiency of Mfrn1 results in a profound hypochromic anemia in vertebrate species. To map the cis-regulatory modules (CRMs) that control expression of the Mfrn genes, we utilized genome-wide chromatin immunoprecipitation (ChiP) datasets for the major erythroid transcription factor GATA-1. We identified the CRMs that faithfully drive the expression of Mfrn1 during blood and heart development and Mfrn2 ubiquitously. Through in vivo analyses of the Mfrn-CRMs in zebrafish and mouse, we demonstrate their functional and evolutionary conservation. Using knockdowns with morpholinos and cell sorting analysis in transgenic zebrafish embryos, we show that GATA-1 directly regulates the expression of Mfrn1. Mutagenesis of individual GATA-1 binding cis elements (GBE) demonstrated that at least two of the three GBE within this CRM are functionally required for GATA-mediated transcription of Mfrn1. Furthermore, ChiP assays demonstrate switching from GATA-2 to GATA-1 at these elements during erythroid maturation. Our results provide new insights into the genetic regulation of mitochondrial function and iron homeostasis and, more generally, illustrate the utility of genome-wide ChiP analysis combined with zebrafish transgenesis for identifying long-range transcriptional enhancers that regulate tissue development.

Iron metabolism and heme synthesis are two key regulators of erythropoiesis. Erythroid differentiation is characterized by the de novo production and activation of numerous enzymes and transporters that are responsible for the synthesis of heme and Fe/S clusters in the mitochondria (2, 42). We previously showed that the anemic zebrafish mutant frascati carries a mutation in the mitochondrial solute carrier, mitoferrin 1 (Mfrn1; Slc25a37), which imports iron into the mitochondria for heme and Fe/S cluster biogenesis in erythroblasts (43). A closely related paralog, mitoferrin 2 (Mfrn2; Slc25a28), functions analogously in nonerythroid tissues and early erythroid progenitors (39, 43). Both Mfrn genes are differentially expressed in a tissue- and developmentally restricted manner. In zebrafish and mouse, Mfrn1 expression is restricted predominantly to hematopoietic tissues, whereas Mfrn2 is ubiquitously expressed (43). Recent studies have demonstrated that Mfrn1 is in part regulated by posttranslational protein stability (39), so that Mfrn1 protein forms an oligomeric complex with Abcb10 to enhance its protein half-life and promote the influx of iron for heme synthesis in erythroblasts (15).

GATA proteins are a family of related transcription factors containing two zinc finger domains: the C-terminal zinc finger is required for DNA binding, while the N-terminal finger stabilizes DNA binding and facilitates physical interaction with their close interacting partners, the FOG (friend-of-GATA) proteins, through direct binding (11, 21). Among the six known members, GATA-1, GATA-2, and GATA-3 are expressed in hematopoietic, reproductive, endocrine, and exocrine cells (8, 12, 31), as well as a specific class of neurons in the central and peripheral nervous system (34, 36). The other GATA factors, GATA-4, -5, and -6, are not expressed exclusively in hematopoietic tissues (30, 35, 40). GATA-1 is highly expressed in mature erythroid cells, where it binds the DNA consensus sequence (T/A)GATA(A/G) in a variety of genes necessary for erythroid maturation (17, 26, 54), including those required for iron acquisition and heme synthesis (44). Therefore, it is likely that GATA-1 participates in regulating mitochondrial iron transport by activating Mfrn1 gene expression.

Critical functions of GATA-1 in erythroid differentiation are achieved through activation and repression of many genes implicated in all steps of erythroid cell maturation (11). Thus, GATA-1 upregulates the erythroid genetic program and suppresses genes involved in the establishment and maintenance.
of early pluripotential hematopoietic progenitors (10). The precise balance between the transcriptional factors GATA-1 and GATA-2 is part of the erythroid cell fate decision (9). The “GATA switch,” where an increase in the GATA-1 level, facilitated by GATA-1 displacement from its binding site, provides an important mechanism of GATA-1 chromatin occupancy for GATA-related activation, as well as its repression (3, 9, 28, 38).

Genome-wide chromatin immunoprecipitation (ChIP) databases for erythroid chromatin occupancy by GATA-1 provide a powerful resource to identify developmentally important target genes and their associated cis elements or cis-regulatory modules (CRMs) (17, 26, 54). However, these ChIP assays and associated bioinformatics predictions require further in vivo functional validation. Zebrafish offer many practical advantages to facilitate the rapid identification of functional cis elements. These advantages include a large number of progeny that mature rapidly (30, 53), transparent embryos to facilitate identification of embryonic tissues, and highly efficient methods of transgenesis (49).

We used transgenesis in zebrafish and mice to identify the transcriptional regulatory elements that regulate expression of the Mfrn genes during vertebrate development. We first identified a cytosine-phosphate-guanosine (CPG) island, encompassing the promoter region in the Mfn2 locus, which recapitulates the endogenous expression pattern for the Mfn2 transcript. We then identified two highly conserved CRMs in the Mfn1 locus, the −20.4-kb CRM and the −37.5-kb CRM, which specifically control the expression of Mfn1 during vertebrate hematopoiesis. The −37.5-kb CRM region assayed in this study directed transgene expression to the zebrafish heart, brain, and blood cells and thus recapitulated many of the salient and conserved aspects of endogenous mfrn1 gene expression. A developmental analysis, using fluorescence reporter transgenes and confocal microscopy, revealed that the Mfn transgenes were restricted in expression to the myocardium of the heart. Finally, using knockdowns, fluorescence-activated cell sorting (FACS), ChIP assays, and mutagenesis of GATA-1 binding cis elements (GBE), we showed that the GATA-1–FOG-1 complex directly regulates the activation of Mfn1 during erythropoiesis.

MATERIALS AND METHODS

Zebrafish transgenic lines and WISH. Wild-type (AB* and Tü), cloche (cloro)11, and transgenic Danio rerio (Danio rerio) were maintained according to standard methods (53). The following transgenic zebrafish lines were used: myocardium-specific Tg(cmlc2:DsRed)7797 (18), endothelium-specific Tg(Bk:mc Cherry-ras)5889, and erythroid-specific Tg(TG1:DsRed)4571. Whole-mount in situ hybridization (WISH) was carried out as described previously (1). All zebrafish studies were conducted under use and approval of the Institutional Animal Care and Use Committee (IACUC) at Children’s Hospital Boston.

Zebrafish anti sense MO injections and RT-PCR. Morpholinos (MOS) against zebrafish GATA-1, FOG-1 (1), and GATA-2 (5′-GAGGACGCTAAGAAATATAAC-3′) were obtained from GeneTools (Philomath, OR); 0.03 to 0.06 pmol MO was injected into embryos at the 1- or 2-cell stage. GATA-2-specific primers were used to perform a semiquantitative reverse transcription (RT)-PCR using total RNA and standard procedures (1).

Identification of promoter-enhancer elements for the Mfn genes. Mouse genomic Mfn1 and Mfn2 sequences were downloaded from the Ensembl Genome Browser (Sanger Centre, United Kingdom) and displayed using the University of California, Santa Cruz (UCSC) genome database. Candidates for CRMs with GBE associated with high ChIP peaks (17, 26, 54) were identified using the Pennsylvania State University (PSU) Genome Browser (http://main.genome-browser.bx.psu.edu). A candidate CRM was selected in the regions with GBE peaks in both ChIP-on-chip DNA microarray (ChIP-chip) and ChIP sequencing (ChIP-seq) for analysis in transgenic zebrafish.

Gatewaysubcloning of the mouse CRMs into expression vectors and expression of transgenes in zebrafish. Gateway-compatible vectors (Invitrogen) were used to clone the expression of the candidate CRMs (49). The molecular and transgenic techniques are detailed in our previous studies (1, 49). Embryos were coinjected at the one-cell stage with 25 pg Tol2 transposase cRNA and 50 pg Tol2-flanked destination green fluorescent protein (GFP) reporter plasmid (49). The expression of the transgenic GFP reporter was visualized at 24 h postfertilization (p.f.) under a fluorescence microscope (Leica MZFLIII).

Mutagenesis of the GATA-1 binding sites. Regions of the −37.5-kb CRM of the mfrn1 gene containing the 3 GATA-1 binding sites were shown in Fig. 1. Mutation of 2 nucleotides at the central core of the GATA sequence (GATA−CTTA) (16, 55) was performed through conventional overlap extension PCR. The resulting PCR products were cloned in the Gateway-compatible vectors (49), and their sequences were confirmed by sequence analysis. The primers used in the overlap extension PCR are listed in Table 1.

Fluorescence-activated flow cytometry. The morpholino targeting GATA-1 was injected into fertilized eggs at the 1- or 2-cell stage collected from Tg(mfrn1:GFP) zebrafish. Single-cell suspensions for flow cytometry from approximately 100 to 200 embryos were prepared from MO-injected and control clutches and sorted as previously described (1).

Mouse transient transgenesis. Pronuclear injections were carried out at the Transgenic/Genetic Targeting Mouse Facility at the University of Utah. Briefly, pronuclei of fertilized eggs were microinjected with linearized DNA, incubated at room temperature, and implanted into a 0.5-day pseudopregnant female. Embryos were dissected in cold 1× phosphate-buffered saline (PBS) at day 8.5 from the day of embryonic implantation (E8.5) and 12.5. GFP emission was immediately acquired on the SXZ12 stereomicroscope (Olympus) equipped with a Bh2-RFL-T3 fluorescent lamp (Olympus). The embryos and yolk sacs were fixed for 20 min (E8.5 yolk sac) to 1 h (E12.5 embryos) in 4% paraformaldehyde (PFA)/PBS at 4°C with rocking. Sections were coveredslipped with Vectashield (Vector Laboratories) and imaged on a BX51 fluorescence microscope (Olympus) or an FV1000 confocal microscope (Olympus). All mouse studies were conducted under the guidance and approval of the IACUC at the University of Utah (Salt Lake City, UT) and the Jackson Laboratory (Bar Harbor, ME).

Staining and histology of mouse embryos. Matings were established between C57BL/6J females (wild-type) and C57BL/6J congenic male mice heterozygous for the Mfrn null allele (45), a gene trap in which the sequence encoding β-galactosidase is inserted downstream of the endogenous Mfrn1 promoter (UCSF BayGenomics mES clone XB454). For whole-mount photography, yolk sacs obtained at E8.5 and embryos obtained at E12.5, E13.5, and E14.5 were fixed briefly and stained for β-galactosidase activity for 48 h (25). All reagents were obtained from major vendors. Abdominal and cardiac cavities were opened to promote contact with the stain. Prior to being imaged (see below), yolk sacs were fixed in 4% paraformaldehyde overnight and transferred through graded ethanol (25). Embryos were dissectioned in PBS and stained with 0.1% (vol/vol) aqueous X-gal solution for 37°C overnight. Sections were photographed using a Nikon SMZ1500 dissecting microscope and a Retigie 2000R digital camera (12 bit) with QCapture software, version 3.0.

For sectioning, embryos stained as described above were fixed overnight in Bouin’s fixative and embedded in paraffin. Digital images of 10-μm sections were collected using a Nikon E600 microscope equipped with Nikon 10×0.3 and 40×0.75 objective lenses. Images were captured with a model 2.21 RT-Color Spot digital camera and Spot software version 4.5.

ChIP assay and real-time PCR. ChIP assays were performed as previously described (1, 54). Friend mouse erythroleukemia (MEL) cells were chemically induced to undergo differentiation by the addition of dimethyl sulfoxide (DMSO) to a final 1.7% (vol/vol) concentration in Dulbecco modified Eagle medium with 10% fetal calf serum, 2% penicillin/streptomycin, and 1% glucose (Gibco). Cells were incubated with 0.5 μmol/L of colcemid (Hoechst) for 30 min prior to harvesting. Nuclei were prepared by hypotonic lysis and centrifuged at 2500 × g for 10 min. ChIPs were performed as described (1). Primers used for ChIP are listed in Table 1.
Statistical analysis. Pooled data were calculated as mean ± standard deviation (SD) with the indicated number of repeats. Pairwise comparison was performed by t test and multiple comparisons by analysis of variance (ANOVA) using the SAS 9.1 software.

RESULTS

Locus-wide GATA-1 ChIP analyses identified Mfrn1 and Mfrn2 transcriptional regulatory regions. In vertebrates, the expression of Mfrn1 is largely restricted to the hematopoietic tissues (43), and GATA-1 regulates most of the genes that define the mature erythroid phenotype (11, 30). To map candidate CRMs that direct the expression of the Mfrn genes in teleosts and mammals, we used genome-wide predictions by ChIP analysis to search for candidate GBE (17, 26, 54). The mouse Mfrn1 locus contains three prominent GBE peaks (Fig. 1, red boxes) for both GATA-1 ChIP-seq (Fig. 1, black boxes) and GATA-1 ChIP-chip (Fig. 1, red peaks), designated 20.4 kb, 34.0 kb, and 37.5 kb (for distances in kb relative to the TSS). Three conserved core sequences for GATA-1 binding sites over the −37.5-kb CRM are shown as pink blocks. The orange numbers indicate the nucleotide positions. The image was generated from the PSU genome bioinformatics site (17).

![FIG. 1. Genome-wide analysis of ChIP-seq and ChIP-chip data predict CRMs. The Mfrn genes are plotted at the top using the base positions in mouse chromosomes (chr) 14 and 19. The coding regions are represented by the blue boxes. The intronic regions are shown as arrowheads representing the direction of transcription. The TSS is indicated by the bent arrow. The G1R-ER4/G1E inputs are represented by black boxes. The positions of the CpG islands are shown as green boxes. The regulatory-potential scores, based on a seven-species sequence alignment, are shown as blue peaks. The murine amplicons used in zebrafish and mouse transgenesis are shown as red (ChIPs) or green (CpG) rectangles above the line, with the positions of the amplicons relative to the TSS (−20.4 kb, −34.0 kb, and −37.5 kb). Three conserved core sequences for GATA-1 binding sites over the −37.5-kb CRM are shown as pink blocks. The orange numbers indicate the nucleotide positions. The image was generated from the PSU genome bioinformatics site (17).](image-url)

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence (5′–3′)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PCR ChIP</td>
<td></td>
</tr>
<tr>
<td>GATA1 HS1-F</td>
<td>TGATGGCTTCTACTAGGCACACG</td>
</tr>
<tr>
<td>GATA1 HS1-R</td>
<td>GGCTTCACCTCCCGAGAAATGAG</td>
</tr>
<tr>
<td>Mfrn1-35.7 kb-F</td>
<td>CTCTAGGGCCCAGGCTAGAT</td>
</tr>
<tr>
<td>Mfrn1-35.7 kb-R</td>
<td>ACACGGCATGCAAGGTTAT</td>
</tr>
<tr>
<td>Mfrn1-20.4 kb-F</td>
<td>TCTTGGTGCTTGATAGGG</td>
</tr>
<tr>
<td>Mfrn1-20.4 kb-R</td>
<td>TGATGTCAGTGCCACAAGGT</td>
</tr>
<tr>
<td>Mutagenic GATA-1 PCR</td>
<td></td>
</tr>
<tr>
<td>GATA1 M1-F</td>
<td>ATGTTTCATATTCTTAGAGGCACACG</td>
</tr>
<tr>
<td>GATA1 M1-R</td>
<td>CAAGCGCAGCCCTTTAAGATAGTAAACGTAG</td>
</tr>
<tr>
<td>GATA1 M2-F</td>
<td>GAACAGCAGTTTAAAGATGTCCTCCAACCC</td>
</tr>
<tr>
<td>GATA1 M2-R</td>
<td>GGAGAGGTAGGTTTAAGCTTCCAGT</td>
</tr>
<tr>
<td>GATA1 M3-F</td>
<td>CCTTCAACGCGCTCGGTGTTTTC</td>
</tr>
<tr>
<td>GATA1 M3-R</td>
<td>CATGCAGAGGTTAAGAGTCTTCCAGT</td>
</tr>
</tbody>
</table>

*F, forward primer; R, reverse primer; HS1, upstream GATA-1 gene enhancer; M1, M2, and M3, GATA-1 mutants.*
transcriptional start site [TSS]). A high regulatory-potential (RP) score (23, 52) was used to detect similarity to sequence patterns in alignments distinctive of regulatory regions in seven vertebrate species (Fig. 1, blue peaks). The RP score is a computational tool to aid in the identification of putative regulatory sites in vertebrate genomes. Unlike search tools based on known transcription factor binding site motifs, RP scores approach the problem from a comparative-genomics perspective. The scores are computed from a comparison of genome-wide alignments of highly conserved DNA sequences from humans and vertebrate species (23).

transcriptional start site [TSS]). A high regulatory-potential (RP) score (23, 52) was used to detect similarity to sequence patterns in alignments distinctive of regulatory regions in seven vertebrate species (Fig. 1, blue peaks). The RP score is a computational tool to aid in the identification of putative regulatory sites in vertebrate genomes. Unlike search tools based on known transcription factor binding site motifs, RP scores approach the problem from a comparative-genomics perspective. The scores are computed from a comparison of genome-wide alignments of highly conserved DNA sequences from humans and vertebrate species (23). A fourth putative GBE peak, localized ~6 kb downstream of the TSS, was not functionally interrogated because of its low RP score. Both Mfrn genes contain CpG-rich regions (Fig. 1, green boxes), which have been associated with promoter activity for expressed genes (4). We examined the −20.4-kb CRM, −34.0-kb CRM, and −37.5-kb CRM sequences and found two or three conserved core GATA-binding motifs (Fig. 1 and data not shown).

Comparative in vivo analysis of promoter regions for Mfrn1 and Mfrn2 genes. Aside from its hematopoietic expression in the yolk sac blood island equivalent (the intermediate cell mass [ICM]), mfrn1 is also expressed in regions of the central nervous system (CNS) in zebrafish (Fig. 2A and data not shown). In contrast, mfrn2 is more ubiquitously expressed in the CNS and somitic myotomes (Fig. 2B and data not shown). Moreover, the mfrn2 transcript is not detected in the developing ICM (Fig. 2A and B). At 48 h p.f., both mfrn1 and mfrn2 show similar patterns of expression in the CNS (data not shown).

To investigate the in vivo function of the CpG-rich regions and the GBE in the respective Mfrn1 and Mfrn2 loci, the corresponding mouse genomic sequences were functionally interrogated in transgenic zebrafish. Briefly, each mouse sequence was inserted using the reporter Gateway system into a Tol2 transposon-based vector containing a GFP reporter cassette for coinjection with Tol2 mRNA as previously described (1). The resulting constructs were injected into one-cell zebrafish embryos and evaluated for GFP reporter expression at 24 h p.f. The CpG island surrounding the promoter region of Mfrn2 gave rise to transgenic fish expressing GFP in the retina, myotome, and CNS that recapitulate the endogenous mfrn2 expression pattern (Fig. 2B). A majority of the injected embryos that survived showed the same consistent expression pattern between 24 and 96 h p.f. (data not shown). The surviving embryos were
screened for germ line integration to generate the stable Tg(Mfrn2:GFP) transgenic fish (Fig. 2D and data not shown). No GFP expression was observed in embryos injected with the corresponding CpG island from Mfrn1, indicating that other transcriptional regulatory elements might be necessary for its expression (data not shown).

**Occupancy by GATA-1 accurately predicts Mfrn1 CRM in zebrafish and mouse.** To test the in vivo functions of the three candidate CRMs in the Mfrn1 locus, we proceeded as described above. Injections of the −37.5-kb CRM and −20.4-kb CRM produce transient transgenic zebrafish expressing GFP in the ICM (Fig. 3 and Table 2). Importantly, the −37.5-kb CRM directed the reporter gene expression in the same tissues in which mfrn1 mRNA is normally expressed (Fig. 2A and C). In contrast, the −20.4-kb CRM resulted in a lower level of GFP expression (Table 2). Injection of the −37.5-kb CRM also produced a high frequency of promiscuous expression in the hatching gland (HG), where the mfrn1 mRNA is not normally expressed (Fig. 3). The GFP expression using a construct that contains the −37.5-kb CRM linked with its CpG island more effectively recapitulated the endogenous mfrn1 expression pattern, excluding the leaky expression in the HG (Fig. 3). Therefore, we generated the stable Tg(Mfrn1:GFP) using the −37.5-kb CRM combined with the endogenous Mfrn1 minimal promoter (CpG island). Henceforth, we refer to the Tg(Mfrn1:GFP) transgenic construct as containing the −37.5-kb CRM and the endogenous Mfrn1 minimal promoter. No consistent or significant expression of the GFP reporter was seen in any of the zebrafish transgenic embryos injected with the −34.0-kb CRM (Table 2). To confirm that the Tg(Mfrn1:GFP) drives the expression in the hematopoietic compartment, double Tg (Mfrn1:GFP)-Tg(GATA-1:DsRed) transgenic fish were generated (Fig. 2E to E'). At 24 h p.f., erythroid progenitor cells in the posterior ICM expressing the red fluorescent protein DsRed also coexpressed GFP (Fig. 2E').

To determine the fidelity and general application of our observations to mammalian transcriptional regulation, we assayed the activities of Mfrn1 CRMs in mouse embryos. We focused our analysis on the −37.5-kb CRM construct, which accounts for the most robust filial expression of endogenous Mfrn1 in teleosts. Briefly, the Mfrn1 CRM construct containing GFP as a reporter was injected into mouse pronuclei. The GFP reporter was analyzed at E8.5 and E12.5 developmental stages and compared to the endogenous mouse Mfrn1 expression. Consistent with our previous observations in the zebrafish, the Mfrn1 −37.5-kb CRM directs GFP expression in the mouse tissues where Mfrn1 is normally expressed (Fig. 4). At E8.5, during primitive hematopoiesis, the Mfrn1 CRM drives GFP expression in the yolk blood islands, the site of primitive erythropoiesis (Fig. 4A and B). At E12.5, the Mfrn1 CRM activates GFP expression in the fetal liver, the site of definitive erythropoiesis (Fig. 4C to F).

We evaluated various CRM combinations to determine their levels of GFP reporter expression. The tandem combination of −34.0-kb CRM plus −37.5-kb CRM or −20.4-kb CRM plus −37.5-kb CRM did not result in any difference in the expression of the GFP reporter in zebrafish embryos compared to

---

**TABLE 2. Functional analysis of Mfrn1 candidate CRMs by GFP transgenesis**

<table>
<thead>
<tr>
<th>CRM (kb)</th>
<th>Stage (h.p.f.)</th>
<th>No. of injected embryos</th>
<th>No. of implanted embryos</th>
<th>No. of surviving embryos</th>
<th>No. of GFP embryos</th>
<th>% (GFP+ implanted embryos)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mouse</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>−20.4 kb</td>
<td>E8.5, E12.5</td>
<td>95, 103</td>
<td>50, 72</td>
<td>1, 0</td>
<td>2, 0</td>
<td></td>
</tr>
<tr>
<td>−37.5 kb</td>
<td>E8.5, E12.5</td>
<td>68, 115</td>
<td>23, 70</td>
<td>3, 10</td>
<td>13, 14</td>
<td></td>
</tr>
<tr>
<td>Zebrafish</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>−20.4 kb</td>
<td>24 h.p.f.</td>
<td>274</td>
<td>156</td>
<td>36, 36</td>
<td>23</td>
<td></td>
</tr>
<tr>
<td>−34.0 kb</td>
<td>24 h.p.f.</td>
<td>118</td>
<td>92</td>
<td>0, 0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>−37.5 kb</td>
<td>24 h.p.f.</td>
<td>578</td>
<td>392</td>
<td>344, 88</td>
<td>88</td>
<td></td>
</tr>
</tbody>
</table>
Yolk sac blood islands

A

B

Mfm1

Tg(Mfm1:GFP)

YS

YS

E8.5

E8.5

Fetal liver

C

D

Mfm1

Tg(Mfm1:GFP)

FL

FL

E12.5

E12.5

FIG. 4. Analysis of the murine CRMs in transgenic mice. (A) The mouse Mfm1 gene is highly expressed in the yolk sac (YS) blood islands. (B, D, and F) Transgenic-mouse embryos carrying the murine −37.5-kb CRM for the Mfm1 gene show strong expression in the yolk sac, as well as the hematopoietic fetal liver (FL), comparable to the endogenous Mfm1 mRNA expression in the embryonic YS (A) and FL (C). (E and F) Higher magnifications of the transversely sectioned fetal liver from panels C and D are shown. The mouse embryonic stages are indicated in the bottom left corners.

−37.5-kb CRM alone in terms of tissue fidelity or intensity (data not shown).

Occupancy by GATA factors at the distal Mfm1 CRMs. To correlate the transgenic in vivo enhancer activity with the in vivo transcriptional-factor occupancy during erythroid matura-

ChIP experiments were performed with characterized antibodies (1, 54) by using β-estradiol-inducible G1ER and MEL cell lines. The G1ER (G1R-ER4) cells are derived from GATA-1-null embryonic cells, which stably express a conditional form of GATA-1 in which the full-length protein is fused to the ligand-binding domain of the estrogen receptor (51). The addition of estradiol to G1ER cells activates GATA-1 and triggers erythroid maturation by nuclear translocation of the fusion protein (51, 52). In undifferentiated erythroid cells, the −37.5-kb CRM is occupied by GATA-2, which functions in hematopoietic stem cell (HSC) and early erythroid precursors (22) (Fig. 5A). In β-estradiol-induced G1ER and DMSO-
treated MEL cells, GATA-1 was significantly enriched at both −20.4-kb and −37.5-kb CRMs in these more mature erythroid cells (Fig. 5A and B). Similarly, FOG-1 was strongly enriched at both CRMs in G1ER and MEL cells induced to undergo erythroid differentiation (Fig. 5A and B). The level of GATA-2 occupancy at both CRMs was very low in MEL cells (data not shown). Similar results for low GATA-2 expression in uninduced and induced MEL cells have been reported (54).

The GATA-1 binding sites are involved in −37.5-kb CRM enhancer-like activity. As shown in Fig. 1, the −37.5-kb CRM contains three GBE, designated GBE1, GBE2, and GBE3 (Fig. 6A). To determine whether these binding sites are func-
tionally responsible for the hematopoietic-cell-specific enhancer activity, the GATA core binding sites were mutagenized (GATA→CTTA) (16, 55). The mutant constructs were interrogated in vivo using microinjections in wild-type embryos at the one-cell developmental stage. The mutations in the three GBE were labeled M1, M2, and M3 (Fig. 6A). The effects of the GBE mutations on the −37.5-kb CRM transcrip-
tional activity were evaluated by expression of the GFP reporter. The embryos injected with either control or mutated constructions were placed in 3 different categories (normal, reduced, and absent), depending on the intensity levels of GFP expression in the ICM (Fig. 6B to B*). The single GBE mutant M2 construct resulted in a marked reduction of GFP expression, indicating GBE2 is a critical regulatory cis element (Fig. 6A, G, and H compared with C and D). The M1 and M3 single mutations resulted in normal GFP reporter expression compared with the wild-type construct (Fig. 6A, E, F, I, and J). The M2 mutation resulted in ~65% reduction in transcriptional activity compared with the wild-type construct, whereas the double M1-M2 and M2-M3 mutations resulted in complete absence of GFP reporter expression (Fig. 6A, K, L, O, and P). The M1-M3 double mutation resulted in 90% absence of GFP expression (Fig. 6A, M, and N). These results suggest that GBE2 is necessary but not sufficient for full transcriptional activity. GBE2 requires cooperative interaction with one of its adjacent GBE (either GBE1 or GBE3) for full GATA-mediated transactivation at the −37.5-kb CRM.

GATA-1 and FOG-1 transcriptionally regulate mfrn1 during early erythropoiesis. In zebrafish, the cells destined for hematopoiesis express GATA-1, FOG-1, and GATA-2 transcription factors in the lateral-plate mesoderm (LPM), which subsequently gives rise to the ICM (1, 27). We used WISH to compare the expression patterns of hematopoietic markers and mfrn1 at early stages of embryogenesis. The zebrafish GATA-2 and FOG-1 are expressed as maternal transcripts (1); in contrast, GATA-1 and mfrn1 are expressed later in develop-
ment when zygotic gene transcription is activated. Between the 15- and 20-somite stages, mfrn1 is coexpressed with GATA-1, GATA-2, FOG-1, bm02, and scl (Fig. 7). These re-

sults, combined with our bioinformatics ChIP-seq and GBE mutagenesis studies shown in Fig. 1, 2, 5, and 6, suggest that mfrn1 expression is dependent on GATA and FOG-1 factors during erythropoiesis.

To investigate whether GATA-1, FOG-1, or GATA-2 reg-

ulates the expression of mfrn1, we depleted zebrafish embryos of these transcription factors by injecting antisense MOs. The MO-injected embryos (morphants) were processed by WISH to reveal the expression of mfrn1 in erythroid progenitor cells
between the 15- and 20-somite stages. Control embryos showed normal expression of mfrn1 (Fig. 7), whereas the embryos injected with either GATA-1 or FOG-1 MO showed markedly reduced mfrn1 expression in the LPM and ICM (Fig. 8A). The absence of mfrn1 expression noted in GATA-1 and FOG-1 morphants was not observed in GATA-2 morphants (Fig. 8A), indicating that the loss of mfrn1 expression is specific to the loss of either GATA-1 or FOG-1 function. These results suggest that the occupancy of GATA-2 at the mfrn1 enhancer elements (Fig. 5A) does not activate its transcription in blood. As a control for full knockdown of GATA-2 in morphant embryos, we used RT-PCR to show that the GATA-2 MO resulted in the production of an aberrantly spliced mRNA, leading to subsequent nonsense-mediated mRNA decay (data not shown). To exclude the possibility that the loss of mfrn1 expression was caused by the absence of hematopoietic progenitors in the GATA-1 morphants, we examined the expression of early hematopoietic markers, scl and lmo2, between the 15- and 20-somite stages. There were no changes in scl and lmo2 expression in the GATA-1 morphants compared with control embryos (Fig. 8A), indicating that GATA-1 specifically regulates mfrn1 expression during erythroblast maturation. These results are consistent with our previous analysis, where loss of the FOG-1 cofactor did not perturb early erythroid formation in the LMP and ICM (1). Our data suggest that GATA-1 and FOG-1 transcription factors activate the expression of Mfrn1 in the developing erythron. Our combined ChiP (Fig. 5) and morpholino (Fig. 8) data approximate normal erythroblast differentiation, where GATA-2 is expressed at relatively high levels in early erythroid progenitors and is displaced by GATA-1 and FOG-1 during erythroid maturation, as schematically shown in Fig. 8D.

GATA-1 regulates the Mfrn1 −37.5-kb CRM enhancer in vivo. By 22 h p.f., the expression of GFP in the Tg(Mfrn1:GFP) embryos is localized to the CNS and erythroid progenitors in the ICM (Fig. 2C and E'). Although GATA-1 MO-injected Tg(Mfrn1:GFP) transgenic embryos show no difference in the GFP reporter signal in the CNS, there is complete absence of GFP expression in the ICM (Fig. 8B). GATA-1 is therefore exclusively required for transcriptional activation of the Mfrn1 −37.5-kb CRM in erythroid progenitors (Fig. 8A and B). We used FACS to more precisely quantify the number of GFP-expressing (GFP+) cells in the GATA-1 morphant using the Tg(Mfrn1:GFP) transgenic line. FACS analysis of uninjected Tg(Mfrn1:GFP) embryos represents the total number of GFP+ cells, including those in the CNS and ICM (Fig. 8C). In contrast, GATA-1 morphants show a drastic reduction in the number GFP+ cells in the ICM (Fig. 8B and C). The FACS results, along with our earlier WISH and GBE mutagenesis data, indicate that the Mfrn1 enhancer is a direct transcriptional target of GATA-1 during erythropoiesis.

Mfrn CRMs regulate expression of the GFP reporter in the developing heart. Both transgenic lines, Tg(Mfrn1:GFP) and Tg(Mfrn2:GFP), drive the expression of GFP in the developing cardiac atrium and ventricle at 48 h p.f. (Fig. 9 and 10). Transverse sections of a 48-h p.f. zebrafish embryo revealed that mfrn1 is expressed in the heart and kidney and at the midbrain-hindbrain boundary (Fig. 9B). During early development, the vertebrate heart is composed of two layers, the outer myocardium and the inner endocardium (Fig. 9C and D). We used a transgenic line that expresses the red fluorescent protein DsRed in the myocardial cell layer, Tg(cmhc2:DsRed) (Fig. 9E'), to generate a double-transgenic line, Tg(Mfrn1:GFP)-Tg(cmhc2:DsRed) (Fig. 9E to E'). Whole-mount confocal flu-
orescence images show that the GFP signal is colocalized in the myocardial cell layer (Fig. 9E/H11630). A similar expression pattern was observed for Tg(Mfrn2:GFP) (Fig. 10). We next examined the Tg(Mfrn1:GFP) expression (Fig. 9F) in a transgenic line that delineates the endothelial cell layer, Tg(flk1:mCherry-ras) (Fig. 9F/H11032) by generating a double-transgenic fish, Tg(Mfrn1:GFP)-Tg(flk1:mCherry-ras) (Fig. 9F/H11033). Neither Mfrn transgenic line directed GFP expression in the endocardial cell layer (Fig. 9F/H11033 and 10D/H11033), shown by the lack of colocalization signals. Because the cloch mutant specifically lacks an endocardial layer in its developing heart (45), we examined the expression of our Mfrn transgenes in this genetic background. Both GFP reporters regulated by the Mfrn CRMs were observed in cloch mutants (Fig. 9G and 10E). Together, these results indicate that the GFP expression regulated by both Mfrn transgenic lines is in the myocardial layer (Fig. 9 and 10).

In E12.5 transgenic mice, the −37.5-kb CRM GFP reporter is expressed in the fetal heart and liver at levels comparable to those in the endogenous mouse Mfrn1 (Fig. 9H and I). Examination of mouse strains carrying a lacZ gene trap marker in the Mfrn1 gene (43) showed β-galactosidase staining in embryos at E12.5 to E14.5 confined to the heart (Fig. 9J to L). Both atrial and ventricular chambers express the lacZ surrogate marker for Mfrn1 initially, although its expression is subsequently restricted to the atria by E14.5 (Fig. 9L).

**DISCUSSION**

The closely related Mfrn genes encode inner mitochondrial solute transporters that are critical for iron assimilation in the genesis of heme and Fe/S clusters. Their fundamental function is highly conserved in the eukaryotic phyla, since the ∆MRS3/4
The regulation and function of the Mfrn transporters in vertebrate genomes, CpG islands typically occur near the TSSs of genes, surrounding their respective promoter regions (4). Histones bound to CpG islands are the sites of specific modifications (6), and they frequently bind ubiquitous transcription factors (13). The CpG regions exhibit particularly accessible chromatin structures (20). For this reason, they play a fundamental role in predicting promoters (5). Both the Mfrn1 and Mfrn2 genes contain CpG islands surrounding their TSSs; however, only the Mfrn2 CpG island directed expression of the GFP reporter in an manner identical to its endogenous transcript in transgenic zebrafish embryos. This observation is consistent with the notion of CpG islands functioning as promoter elements for housekeeping genes (4). In contrast, the Mfrn1 CpG island failed to direct expression of the GFP reporter to hematopoietic tissues. Similarly, most minimal GATA-regulated promoters (or larger regions) are themselves not sufficient to direct expression in transgenic assays, and they often require enhancers to augment their transcriptional activities (37).

We used genome-wide predictions by chromatin occupancy (17, 26, 54) to identify distal candidate enhancer CRMs with preserved GBE in the mouse Mfrn1 locus. Using this strategy, we found two genomic segments, the 20.4-kb CRM and 37.5-kb CRM, that directed GFP expression in developing erythroid cells during zebrafish and mouse embryogenesis. Whereas the 37.5-kb CRM was sufficient to account for the erythroid expression of Mfrn1, the CpG island surrounding the TSS was required to confer specificity and repress promiscuous expression in the HG. The 37.5-kb CRM murine genomic DNA fragment also faithfully directed expression of the GFP reporter in the cardiac myocardial layer. Our analysis represents the first demonstration that noncoding sequences with preserved GBE in the Mfrn1 locus can direct expression in erythroid and myocardial populations.

In order to evaluate the fidelity of our findings in zebrafish, we tested both Mfrn1 CRMs using transgenesis in mouse embryos. In zebrafish and mice, the 37.5-kb CRM directed the reporter GFP in all tissues where the endogenous Mfrn1 transcripts are normally expressed, whereas the 20.4-kb CRM displayed less consistent GFP expression. The data generated in both species were highly concordant and suggest that transgenesis in zebrafish serves as a rapid, robust system to validate regulatory elements from mammalian species. The subset of
GATA-1-occupied sites active as enhancers shows strong association with the evolutionary preservation for the binding-site motif across mammalian lineages. While the genome-wide ChIP data permit the identification of candidate DNA segments occupied by GATA-1 (17, 26, 54), the use of zebrafish transgenesis allows faster functional validation. We initially evaluated our CRM constructions for enhancer activity in K562 human erythroleukemia cells, and the addition of the 37.5-kb CRM caused only a modest increase in reporter luciferase activity above that of a control vector containing only the -globin promoter (data not shown). This result indicated that the context-dependent functions of the CRM could be more reliably evaluated in vivo using transgenic models, validating the premise of our studies. Similar cautions for using cell lines to functionally interrogate CRMs have been noted by other investigators (7, 17, 48, 50).

The mutational analyses of three GBE in the 37.5-kb CRM demonstrated that mutation of the GBE2 site is necessary but not sufficient for full enhancer activity. We also showed that erythroid-specific transcriptional regulation was mediated by at least two GATA-1 binding sites, namely, GBE2 with either one of its adjacent GBE, in the central core of the 37.5-kb CRM. The observation that the three GBE are conserved in mammals underscores the importance of this regulation (Fig. 1). The consensus GBE3 is fully conserved in rat and mouse, while GBE1 and GBE2 are fully conserved in human, dog, rat, and mouse. These functional studies in transgenic zebrafish suggest that the GBE within the 37.5-kb CRM regulate expression during erythroid maturation.

Our data with morphant embryos indicate that activation of mfrn1 expression is directly dependent on GATA-1 and FOG-1 activity (Fig. 8A). Previous experiments have shown that GATA-1 occupancy of chromatin sites bound by GATA-2 is facilitated by FOG-1 (32, 38). Our ChIP analysis demonstrated that FOG-1 occupies the 37.5-kb CRM in the same regions bound by GATA factors, consistent with FOG-1 facilitating the displacement of GATA-2 by GATA-1 during erythroid maturation (9, 28, 38). Based on these data, we propose...
a model for erythroid Mfrn1 expression in which at early
developmental stages the maternal GATA-2 factor binds to the
-20.4-kb and -37.5-kb CRMs without activating Mfrn1 tran-
scription. Later, during zygotic transition, the activation of
GATA-1 expression, facilitated by FOG-1, induces an ex-
change of nuclear factors at both CRMs. GATA-2 is displaced
by the GATA-1–FOG-1 complex, resulting in the activation of
Mfrn1 transcription (Fig. 8D). This sequence of events approx-
imates previous in vitro observations in which GATA-2 is ex-
pressed at relatively high levels in early erythropoiesis and is
gradually replaced by GATA-1 during late erythropoiesis (52).
This is consistent with models in which GATA factors nucleate
the binding of various protein complexes and changes in pro-
tein/DNA binding ultimately lead to a rapid increase in tran-
scription at the α-globin promoter (3).

The expression of the GFP reporter by the Mfrn1 distal
CRM in the heart and CNS is not surprising, since other
GATA factors, in combination with their FOG cofactors, could
regulate expression in those tissues. In vertebrates, the GATA
family members GATA-4, GATA-3, and GATA-6 are predom-
GATA-2 binding domains at the protein level (30, 35, 40). In addition, /H11011 morphogenesis, and they share intronically expressed during heart development to regulate cardiac /H11022

**FIG. 10.** The transgenic Tg(Mfrn2:GFP) expresses the GFP reporter in the myocardium. (A and B) Lateral views of zebrafish embryos at 48 h p.f. (A) Schematic representation of the zebrafish embryo (http://www.zfin.org). (B) WISH of mfrn2 in sectioned embryos showing expression in the heart (red arrow). (C and C') Confoocal fluorescence images showing the transgenic Tg(Mfrn2:GFP) expression in the heart (green) (C) and the transgenic Tg(cmclc2:DsRed) marker in the myocardium (red) (C'). (C' and C'') Merged confocal images of a double-transgenic Tg(Mfrn2: GFP):Tg(cmclc2:DsRed) embryo reveal colocaled expression of Mfrn2 in the myocardial layer (yellow). (D and D') Confoocal images of the transgenic Tg(Mfrn2:GFP) (green) (D) and the endocardium-specific transgenic Tg(flk1:mCherry) (red) (D'). (D') Merged confocal images in a double-transgenic Tg(Mfrn2:GFP):Tg(flk1:mCherry) embryo indicate Mfrn2 is not expressed in the endothelial-cell layer. (E) Confoocal microscopy of transgenic Tg(Mfrn2:GFP) in the endocardium-deficient mutant (clo) show GFP expression, supporting the idea that Mfrn2 is specifically expressed in the myocardium.

stantly expressed during heart development to regulate cardiac morphogenesis, and they share ~85% similarity in their DNA-binding domains at the protein level (30, 35, 40). In addition, GATA-2 and GATA-3 are expressed during early stages of CNS development, indicating that these factors might collaborate during neural-tissue differentiation (34, 36). Using ChIP assays, we demonstrated that the Mfrn1 CRM is bound by GATA-2. Thus, GATA-2 could bind to the Mfrn1 enhancer to potentially activate its expression in the CNS. Similar to the “cardiac” and “neuronal” GATA factors, their FOG-1–FOG-2 partners are also expressed in both tissues, and their loss of function contributes to several cardiac and neurogenic defects (11, 12). Finally, in the Mfrn1 CRM sequence, we also identified binding motifs for the helix-loop-helix transcriptional factors, MyoD, myogenin, and Myf-5, which are critical for specification and differentiation of muscle cells (41). On the basis of these observations, we propose that those motifs are key elements for driving Mfrn1 CRM myotomal expression. Together, our results demonstrate that analysis of mouse genomic element in zebrafish is suitable for rapid identification of long-range cis-regulatory elements that regulate tissue-specific gene expression.

**ACKNOWLEDGMENTS**

We thank Christian Lawrence, Jason Best, Adam Cardullo, and Paul McBride for help with zebrafish animal husbandry; Jessica Wagner for help with confocal microscopy; Kathleen Solitis for screening cloche and Tg(Mfrn1:GFP) zebrafish; Shuo Lin for the Tg(GATA-1:GFP) transgenic line; David Traver for the Tg(GATA-1:DsRed) transgenic line; and Susan Tamowski for the pronuclear injections. We thank Iman Schultz, Heidi Anderson, Adrienne Kolpak, and John Kanki for reviewing the manuscript. Confocal fluorescence microscopy was performed at the Harvard Digestive Disease Center Imaging Facility (Children’s Hospital Boston).

This work was supported by grants from the American Heart Association (J.D.C.) and the March of Dimes Foundation (B.H.P.) and by National Institutes of Health grants K08 HL074891 (N.C.C.), U01 HL066611 (L.L.P.), R01 DK052380 (J.K.), R01 DK070838 (B.H.P.), P01 HL032262 (A.B.C. and B.H.P.), and P30 DK072437 (J.K. and B.H.P.).

We declare no financial conflict of interest.

**REFERENCES**


