A novel regulatory element between the human FGA and FGG genes

Richard J. Fish1; Marguerite Neerman-Arbez1,2

1Department of Genetic Medicine and Development, University of Geneva Medical Centre, Geneva, Switzerland; 2Service of Angiology and Haemostasis, University Hospital and Faculty of Medicine, Geneva, Switzerland

Summary

High circulating fibrinogen levels correlate with cardiovascular disease (CVD) risk. Fibrinogen levels vary between people and also change in response to physiological and environmental stimuli. A modest proportion of the variation in fibrinogen levels can be explained by genotype, inferring that variation in genomic sequences that regulate the fibrinogen genes (FGA, FGB and FGG) may affect hepatic fibrinogen production and perhaps CVD risk. We previously identified a conserved liver enhancer in the fibrinogen gene cluster (CNC12), between FGB and FGA. Genome-wide Chromatin immunoprecipitation-sequencing (ChIP-seq) demonstrated that transcription factors which bind fibrinogen gene promoters also interact with CNC12, as well as two potential fibrinogen enhancers (PFE), between FGA and FGG. Here we show that one of the PFE sequences has potent hepatocyte enhancer activity.

Using a luciferase reporter gene system, we found that PFE2 enhances minimal promoter- and FGA promoter-driven gene expression in hepatoma cells, regardless of its orientation with respect to the promoters. A region within PFE2 bears a short series of conserved nucleotides which maintain enhancer activity without flanking sequence. We also demonstrate that PFE2 is a liver enhancer in vivo, driving enhanced green fluorescent protein expression in transgenic zebrafish larval livers. Our study shows that combining public domain ChIP-seq data with in vitro and in vivo functional tests can identify novel fibrinogen gene cluster regulatory sequences. Variation in such elements could affect fibrinogen production and influence CVD risk.

Keywords

Fibrinogen, gene regulation, regulatory element, enhancer

Introduction

Elevated circulating fibrinogen has been associated with cardiovascular disease (CVD) risk across a variety of study populations (1). Fibrinogen levels are influenced by environmental and genetic components, with the latter contributing 20–50% of the variability. The central role of fibrinogen as the precursor to fibrin, its ability to cross link platelets, and its effects on blood viscosity suggest a fibrinogen concentration-dependent effect on CVD by promoting clotting and pre-disposing to thrombotic events (2). However, proving a causal role of fibrinogen in disease risk is complicated by its variable concentration in individuals and its increase in response to inflammatory stimuli, which in turn influences coagulation (3). Fibrinogen levels may be a marker of CVD-associated inflammation rather than a direct causal factor in CVD events. A randomised control trial with an appropriate fibrinogen-lowering agent may resolve this issue, but no such treatment is available at present. Recently the role of elevated fibrinogen in promoting thrombosis and increasing clot resistance to fibrinolysis was assessed in a mouse model system and provided experimental evidence for a causal role of fibrinogen levels in both mechanisms (4, 5).

The three fibrinogen polypeptide chains, Aα, Bβ and γ, that each contribute two copies to the functional fibrinogen hexamer, are encoded by a compact 50 kb three-gene cluster (FGB-FGA-FGG) on human chromosome 4 and are expressed primarily in hepatocytes. Coordinate regulation of the three genes has been described (6), and most likely relies on a combination of common transcription factors binding to gene promoters and enhancers (7) and post-transcriptional mechanisms (8) (and our review in this issue [29]). Regulatory polymorphisms in the FGB promoter have been associated with fibrinogen levels (9), but explain only a small percentage of its variation. Other variants influencing fibrinogen levels most likely reside within 1) fibrinogen regulatory sequences 2) pathways that influence fibrinogen production and 3) uncharacterised gene-environment interactions. Genomic sequences in the first category, excluding proximal gene promoters, include non-coding cis regulatory elements; distal enhancers and insulators. Identification of such sequences traditionally relied on cloning and characterising sequences near to genes and identifica-
tion of consensus sequences for transcription factor binding. However, recent advances using high-throughput technologies facilitate the search for regulatory sequences by cataloguing genomewide binding sites for chromatin-interacting factors and modifications, in appropriate cell types.

Inter-species whole genome comparisons have revealed functional enhancer sequences by conservation (10, 11) and by descriptions of transcription factor binding sites using ChIP-seq (12). In particular, the binding of p300, a transcriptional co-activator and histone acetyltransferase, marks enhancer sequences that drive tissue-specific expression patterns when tested in vivo. Similarly, using mammalian sequence conservation, our previous study focusing on the fibrinogen locus enabled the identification of liver enhancer CNC12, residing between the FGB and FGA genes (13). However, descriptions of sequence conservation and synteny are unlikely to uncover all possible regulatory sequences for the human fibrinogen genes. Changes in regulatory element sequence, number and position relative to the genes they influence, over evolutionary time, are likely to be important sources of gene expression diversity between species (14, 15).

Here using public domain transcription factor and histone modification ChIP-seq data, and a combination of cell-based in vitro and transgenic zebrafish in vivo functional tests, we report the identification and functional validation of a novel human fibrinogen locus regulatory sequence.

Materials and methods

Bioinformatics and ChIP-seq data

Human genome representations were obtained using the UCSC Genome Browser (16) on genome assemblies produced by the Genome Reference Consortium. ChIP-seq and histone modification data are from the ENCODE project (17). The ChIP-seq data from HepG2 cells represented in Figure 1A and Supplementary Figure 1A (available online at www.thrombosis-online.com) were contributed to the ENCODE project by the laboratories of Richard Myers, Hudson Alpha Institute for Biotechnology (p300, HNF4A, RXRA, FOXA1, FOXA2, SP1, HDAC2, TCF12, HNF4G and HEY1) and Michael Snyder, Stanford University (CEBPB and TCF4). Histone modification data represented in Supplementary Figure 1B (available online at www.thrombosis-online.com) was contributed to the ENCODE project by the laboratory of Bradley Bernstein at the Broad Institute. Conservation of genomic sequences in mammals is represented in Supplementary Figure 1A and B (available online at www.thrombosis-online.com). Peaks are generated automatically in the UCSC genome browser and are based on genome alignments of placental mammals using MULTIZ (18) and conservation scores using phastCons (19). The genomes aligned to generate these peaks differ modestly between NCBI36/Hg18 and GRCh37/Hg19 human genome assemblies, details are available at the UCSC genome browser website (http://genome.ucsc.edu/).

Cells

HEK-293T cells, HuH7 cells and HepG2 cells were maintained in DMEM or EMEM (HepG2) supplemented with 10% heat-inactivated fetal bovine serum, 100 U/ml penicillin and 100 μg/ml streptomycin (all Invitrogen, Carlsbad, CA, USA).

Luciferase gene reporter plasmid constructions

Regulatory sequences were amplified from human or mouse genomic DNA by PCR and cloned into pCRII TOPO (Invitrogen). Each amplified product was sequenced in pCRII TOPO, and then after sub-cloning to pGL4.10. For pGL4.10 minP, a 94-base pair (bp) mouse c-Fos (Fos) minimal promoter was inserted between the Xhol and BgIII sites. For pGL4.10 FGAP, the human FGA promoter used spans 500 bp upstream of the FGA initiator codon. It was cloned between the pGL4.10 Xhol/HindIII sites. PFE sequences were then cloned in both orientations into the Xhol site of these plasmids. The firefly luciferase (ff luc) reporter gene expression cassettes are described graphically in each figure, and the oligonucleotides used in Supplementary Table 1 (available online at www.thrombosis-online.com). Forward constructs refer to PFE regulatory sequence insertions with the genomic FGB+ strand (centromeric to telomeric in this case) in the same orientation as the luciferase open reading frame, and reverse constructs with the regulatory sequence inverted.

Luciferase reporter gene assay

Cell culture, transfection and luciferase-based gene reporter assays were performed essentially as described previously (8, 13). The pGL4.10 firefly luciferase plasmid (Promega, Madison, WI, USA) was used as the parent (empty) vector for cloning and testing of genomic sequences described. HuH7, HepG2 or HEK-293T cells were seeded in opaque 96-well plates at 10^4 cells per well, one day prior to transfection. pGL4.10-based plasmid DNA was transfected at 167 ng per well with 80 ng per well of pGL4.74 transfection control Renilla luciferase plasmid (pGL4.74), using the FUGENE HD transfection reagent, using the manufacturer’s instructions as a guide (Roche, Basel, Switzerland). Two to four technical replicates were used for each transfection on each plate. A separate transfection mix was used for each plate and considered as an individual experiment with the mean of the technical replicates as the value used. The Dual-Glo luciferase reporter assay (Promega) was used to measure pGL4.10-derived firefly luciferase activity, and pGL4.74-derived renilla luciferase activity as a transfection control, for each well according to the manufacturer’s instructions. Luminescence was measured using a Perkin Elmer Victor3 counter (Perkin Elmer, Waltham, MA, USA), a luminescence detection instrument for 96-well plate assays. Average background luminescence from wells that received the transfection conditions but no
Luciferase plasmids were subtracted from all other data for the firefly and renilla luciferase activity readings. The ratio of this baseline-subtracted firefly/renilla signal was calculated for each well, and values normalised to control condition ratios set at 1 (empty vector or minP vector, depending on the experiment).

**Zebrafish**

AB strain zebrafish were maintained as described previously (20). Experimentation was authorised by the local veterinary authority.

**Tol2 system plasmids**

Tol2 transposon system reagents were obtained from Koichi Kawakami (National Institute of Genetics, Shizuoka, Japan). A XhoI/BamHI fragment from the parent gene transfer vector pT2KXIGΔin was excised and replaced by either the 94 bp mouse c-Fos minimal promoter or a 129 bp mouse β-globin minimal promoter, flanked by compatible restriction sites. The c-Fos sequence was the same used in the cell-based luciferase assays. The XhoI site was then used in each minimal promoter vector for bi-directional cloning of PFE2 or PFE2 171.

---

**Figure 1:** A novel regulatory element between *FGA* and *FGG*. We investigated the regulatory potential of two sequences between human *FGA* and *FGG*: A) The fibrinogen gene cluster with the transcriptional orientation denoted with arrows that span the proximal promoters of each gene. Each PFE sequence shares transcription factor ChIP-seq marks with CNC12, a liver enhancer between *FGB* and *FGA* (13), as well as with the fibrinogen promoter regions. Selected factor binding sites are represented graphically with coloured circles. The data represented was obtained from genomewide ChIP-seq on liver-derived tumour cells (HepG2 cells) as part of the ENCODE consortium. Regulatory sequence activity was assessed using a luciferase-based gene reporter assay in human HEK-293T embryonic kidney and HuH7 hepatoma cells. Firefly luciferase (ff luc) reporter gene expression cassettes are represented to the left of each bar chart. B) The *FGA* promoter is inactive in 293T cells (B, upper graphic, see inset with zoomed x-axis for detail), but over 100-fold more active than the empty luciferase vector in HuH7 cells (B, lower graphic), data were normalised to the empty vector values. C) Enhancer activity of PFE1 and PFE2 was measured in 293T cells and HuH7 cells when inserted in front of a minimal mouse c-Fos promoter (minP), with the PFE sequences in both orientations (F or R). Data were normalised to minP values. When assessed using an unpaired t-test all data were significantly different to the minimal promoter (minP) values in both cell types (p<0.01), except for PFE2F in 293T cells (p=0.69) and PFE1R in HuH7 cells (p=0.056). Data shows mean plus SEM, n=5 to 6 for B, n=6 to 17 for C.
**Transgenic zebrafish gene reporter assay**

The larval zebrafish gene reporter assay has been described previously (13). One- to two-cell stage embryos were micro-injected into the yolk sac with approximately 1 nl of a mixture of 25 pg/nl pT2KXIGΔinin-based reporter plasmid, 35 pg/nl of capped Tol2 transposase mRNA and 0.05% Phenol Red in 1x Danieau buffer (58 mM NaCl, 0.7 mM KCl, 0.4 mM MgSO4, 0.6 mM Ca(NO3)2, 5.0 mM HEPES pH 7.6). Embryos were raised to six days post fertilisation (dpf) and imaged or scored for green fluorescence in the liver under anesthesia using ethyl 3-aminobenzoate methanesulfonate salt (150 μg/ml) in E3 medium (5 mM NaCl, 0.17 mM KCl, 0.33 mM CaCl2 and 0.33 mM MgSO4). Images were acquired using a Leica MZ16FA stereomicroscope equipped with a Leica DFC340FX black and white digital camera at room temperature (Leica, Wetzlar, Germany). Leica LAF v2.1 software was used for imaging. Representative left–side images were obtained in bright field and fluorescence illumination and merged using the Leica LAF software. A 26x magnification with a 2.6x optical zoom and a 1.0x planapochromatic objective with a numerical aperture of 0.06 was used. Images acquired with Leica software were cropped in Adobe Photoshop CS5 without image modification or enhancement, and figures assembled in Adobe Illustrator CS5. For quantification, larvae were scored as liver EGFP positive if a single green fluorescent hepatocyte could be discerned from the tissue background when larvae were rotated under fluorescence illumination.

**Results**

**A liver enhancer element between the human FGA and FGG genes**

Regulatory DNA sequences in the fibrinogen gene cluster include the 5’ proximal promoters (7) of the fibrinogen genes and a conserved liver enhancer sequence (CNC12) between FGB and FGA (13). Recently the ENCODE consortium (17) released genomewide ChIP-seq data for multiple transcription factor binding regions and chromatin marks, in a variety of cell types. In HepG2 hepatoma cells CNC12 is enriched by immunoprecipitation with antibodies to multiple transcription factors, including those found associated with proximal fibrinogen promoter regions. Two short sequences residing between FGA and FGG are also tagged by the same group of factors, suggesting they may also serve as fibrinogen gene cluster enhancers. We named these sequences potential fibrinogen enhancers (PFE1 and PFE2, see schematic representation in Fig. 1A and Suppl. Fig. 1, available online at www.thrombosis-online.com). The ability of PFE1 and PFE2 to enhance gene expression was tested using a firefly luciferase reporter gene. To assess whether our assay detects liver active sequences we used the human FGA promoter as a test sequence in HuH7 hepatoma cells that express fibrinogen and HEK-293T cells (293T), embryonic kidney-derived cells, which do not. The human FGA promoter has no activity in 293T cells but promotes about 100-fold higher luciferase activity in HuH7 cells, compared to an empty vector (Fig. 1B).

![Figure 2: Delineating the functional enhancer region of PFE2](image-url)
Since orientation-independent activity is a criterion for enhancers, PFE1 and PFE2 were cloned in both orientations in front of a minimal promoter and the firefly luciferase gene. In 293T cells, PFE1 showed modest enhancer activity in both orientations, whereas PFE2 only enhanced in one orientation, again weakly. PFE2 showed direction-independent activity in HuH7 cells, consistent with liver enhancer activity, but PFE1 did not give significant enhancer activity in both orientations (Fig. 1C), and was not studied further. Enhancer activity for PFE2 on the same minimal promoter was also measured in HepG2 hepatoma cells (Suppl. Fig. 2A, available online at www.thrombosis-online.com).

**Chromatin marks and short conserved sequences delineate the active PFE2 enhancer element**

Interestingly, PFE2 did not meet the criteria to be tested for enhancer activity in our previous study (13). However, a short region of the 877 base pair PFE2 sequence does contain some mammalian conservation, and the liver-derived cell transcription factor ChIP-seq marks all overlap with this region (Suppl. Fig. 1A, available online at www.thrombosis-online.com). In addition, two histone modifications known to be associated with gene regulatory sequences also define this part of PFE2 in the ENCODE data set. Tri-methylation of histone 3 lysine 4 and the acetylation of histone 3 lysine 27 flank the conserved region of PFE2 (Suppl. Fig. 1B, available online at www.thrombosis-online.com). These "histone marks" appear as evenly-spaced peaks in sequences around PFE2, and are also detected at the fibrinogen gene promoters and CNC12 in a similar fashion (the FGA promoter region is shown in Suppl. Fig. 1B, available online at www.thrombosis-online.com). This may represent the periodicity of specifically marked nucleosomes around regulatory sequences.

To determine whether this region of PFE2 contains the active enhancer sequence we made further enhancer assays in HuH7 and 293T cells. Three separate segments of PFE2 were tested individually; a central 171 base pair conserved region with ChIP-seq transcription factor binding sites and histone marks (PFE2 171) and the two flanking sequences of 270 base pairs (PFE2 270) and 436 base pairs (PFE2 436). As shown in Figure 2, compared to a minimal promoter, PFE2 270, PFE2 171 and PFE2 436 showed no more than two-fold enhancer activity in 293T cells in either orientation. While this is also the case for PFE2 270 and PFE2 436 in HuH7 cells, PFE 171 gave a striking enhancement of luciferase activity in the liver-derived HuH7 cells, with 26.6– and 5.8-fold enhancements for the forward and reverse orientation reporter gene constructions. This demonstrates that the vast majority of the PFE2 liver cell enhancer activity is within the PFE2 171 sequence, which is bound by multiple transcription factors and subject to regulatory histone modifications.

**PFE2 enhances activity of a fibrinogen promoter**

With liver cell enhancer activity established using a proxy minimal promoter, we tested the ability of PFE2 and PFE2 171 to enhance the activity of the human FGA promoter in HuH7 cells. PFE2 and PFE2 171 potently enhance the activity of a 500 nucleotide FGA promoter region in the luciferase assay, compared to the promoter alone (Fig. 3). PFE2 also enhances the activity of the FGA promoter in HepG2 cells (Suppl. Fig. 2B, available online at www.thrombosis-online.com).

**PFE2 drives liver transgene expression in vivo**

To test the ability of PFE2 to drive liver gene expression in vivo we used a ToI2 transposon-based gene reporter in transgenic zebrafish larvae. PFE2 and PFE2 171 were cloned bi-directionally upstream of either the mouse c-Fos minimal promoter, as in the luciferase assay, or a minimal mouse β-globin promoter, with each minimal
Figure 4: PFE2 has liver enhancer activity in vivo. To assess the ability of PFE2 to drive expression of a transgene in the liver in vivo we used transgenic larval zebrafish. PFE2 and PFE2 171 sequences were inserted in both orientations (F/R) upstream of minimal promoters in a Tol2 transposon gene transfer plasmid. Minimal promoters were derived from either the mouse c-Fos gene (Fos), as used for the in vitro luciferase assay, or from the mouse β-globin gene. A) Gene transfer cassettes with or without the PFE2 sequence in red, the minimal promoter (minP) and the EGFP gene with a polyadenylation signal sequence (pA), flanked by Tol2 transposase recognition sequences (Tol2). B) Zebrafish larvae were monitored for green fluorescence at 6 dpf. Images are merged bright-field and fluorescence illumination. The top image shows a non-injected (non-inj) larva, below which are larvae injected with Tol2 plasmid constructs denoted to the right of each. A second non-injected larva is shown above the c-Fos minP constructs because the four lower images were acquired with 800 ms exposure time whereas 400 ms was used for the upper four. The approximate visible liver area is outlined with the same white dotted shape in each image. The fluorescence caudal to the liver in each image is due to autofluorescence of the remaining embryonic yolk sac and developing gut. The scale bar in the upper image applies throughout and represents 250 μm. C) Quantification of liver enhancer activity is represented for transgenic zebrafish larvae at 6 dpf. The percentage of larvae with green fluorescent liver cells is represented by the green part of each bar, green fluorescence-negative livers by white. The number of injected larvae assessed (n) was between 28 and 83 for each transgenic type, with a mean of 51. The left and right graphs show data for the c-Fos and β-globin minimal promoter-containing constructs, respectively. Minimal promoters alone (c-Fos minP and β-globin minP) were used as controls.
Discussion

This study has identified a novel liver enhancer in the human fibrinogen gene cluster. Our in vitro results in human cell lines and in vivo data using zebrafish larvae demonstrate gene regulatory activity for a short sequence between the human FGA and FGG genes. This element, PFE2, shares chromatin regulatory marks with the CNC12 enhancer and the fibrinogen gene promoters and gives balance to the positioning of regulatory elements for the fibrinogen gene cluster; proximal 5′ promoters and liver-active inter-gene enhancers. As with previously reported studies assessing largely developmental enhancers (11,12), the identification of regulatory elements by sequence conservation and then by chromatin marks has unveiled functional parts of the fibrinogen gene cluster.

PFE2 enhances the expression of a transfected luciferase reporter gene in liver-derived cells, regardless of its orientation with respect to a promoter. A short region of the sequence with some mammalian conservation is responsible for the majority of this activity (PFE2 171) and this region is marked by multiple appropriate liver cell transcription factors. The ENCODE transcription factor ChiP-seq sites that we have used as a guide for our study are sequences that immunoprecipitate with antibodies to transcription factors or co-regulator proteins. However, this does not prove definitively that they are binding directly to the marked sequences because the immunoprecipitation technique could cross link sequences in proximity to transcription factor bound sequences, such as promoters, by chromatin looping (21,22). We present, explain the difference in enhancer potency between the forward or reverse configurations of PFE2 or PFE2 171 measured in hepatoma cells. Differences may relate to the artificial nature of our proxy assay using plasmid constructions in transfected cells, compared to the physiological conformation of PFE2 in chromatin.

When introduced into zebrafish embryos PFE2 drives transgene expression in the larval liver, with activity maintained completely in PFE2 171. The PFE2 sequence is not conserved in the zebrafish, at least it is not found by sequence alignment using standard tools. This is not surprising as there is no equivalent of the human FGA to FGG intergenic sequence in the zebrafish (20). As with CNC12 (13), PFE2 remains active in the liver of the zebrafish despite this lack of conservation, implying that the interaction sequences for factors binding to PFE2 are sufficiently conserved to attract binding of factors which are adequately conserved to bind them.

CNC12 and PFE2 are located near the middle of the FGB-FGA and FGA-FGG intergenic sequences, respectively. Although this positions an enhancer element in both fibrinogen locus intergenic regions (Fig. 1A) it remains unclear how the 50 kilobase fibrinogen gene cluster is arranged as an active transcriptional unit in fibrinogen-expressing or indeed in non-expressing cells. Fuller and Zhang proposed a possible looping-out of the fibrinogen gene cluster from its local environment (7), thereby improving accessibility of factors and chromatin modifications to the three genes. Such a configuration would enable CNC12 and PFE2 to act as local concentrators of appropriate transcription factors and co-activators, factors such as HNF4alpha and p300. In the light of recent advances in mapping and understanding chromatin interactions (23) it may soon be possible to detect interactions of promoters and enhancers in and around the fibrinogen gene cluster in fibrinogen-expressing cells.

Our study identifies a new sequence in which variation could impact fibrinogen expression levels and, when combined with other risk alleles, perhaps influence CVD risk. Polymorphisms linked to circulating fibrinogen levels have not been detected in CNC12 or PFE2 by genome-wide association (24); however, this does not exclude a minor contribution of such variants, or very rare alleles affecting these elements, that are below detection in the currently described population studies. Identifying novel elements regulating CVD-associated proteins, proteins such as fibrinogen and other haemostasis-related factors (25–28), will help in the search for a more complete picture of how gene regulatory mechanisms can play a role in CVD risk.

Acknowledgements
The authors thank Corinne di Sanza for expert technical assistance, Koichi Kawakami for T02 reagents and Owen Tamplin and Christian Mosimann for advice on minimal promoters.

Conflicts of interest
None declared.

References


