Co-expression of FBN1 with mesenchyme-specific genes in mouse cell lines: implications for phenotypic variability in Marfan syndrome

Kim M Summers*,1, Sobia Raza1, Erik van Nimwegen2, Thomas C Freeman1 and David A Hume1

Mutations in the human FBN1 gene cause Marfan syndrome, a complex disease affecting connective tissues but with a highly variable phenotype. To identify genes that might participate in epistatic interactions with FBN1, and could therefore explain the observed phenotypic variability, we have looked for genes that are co-expressed with Fbn1 in the mouse. Microarray expression data derived from a range of primary mouse cells and cell lines were analysed using the network analysis tool BioLayout Express3D. A cluster of 205 genes, including Fbn1, were selectively expressed by mouse cell lines of different mesenchymal lineages and by mouse primary mesenchymal cells (preadipocytes, myoblasts, fibroblasts, osteoblasts). Promoter analysis of this gene set identified several candidate transcriptional regulators. Genes within this co-expressed cluster are candidate genetic modifiers for Marfan syndrome and for other connective tissue diseases.

Keywords: Fibrillin-1; Marfan syndrome; mesenchyme; connective tissue; extracellular matrix

INTRODUCTION

Marfan syndrome (OMIM 154700) is an autosomal dominant condition that affects connective tissue.1-4 Individuals show overgrowth of the long bones, lack of adipose and muscle tissue, and abnormalities of the eyes and skin. The major cause of morbidity and mortality is dilatation and dissection of the ascending aorta. Marfan syndrome is usually associated with mutations in the FBN1 gene (OMIM 134797), encoding the microfibrillar protein, fibrillin-1.5,6 A bovine Marfan-like syndrome (OMIA 1204) is also due to a mutation in FBN1. Homozygous mice that lack a functional Fbn1 gene have some manifestations similar to Marfan syndrome in humans, although the heterozygous phenotype is mild.5 In addition, a transgenic mouse line carrying a mutation known to cause severe disease in humans has a dose-dependent phenotype showing aspects of Marfan syndrome.7 A homozygous lethal natural mutation involving duplication of exons 17–40 of the mouse Fbn1 gene8 causes the tight skin (Tsk) phenotype in which heterozygotes have abnormalities of skin, viscera, lungs, cartilage, bone, heart and tendons,9 with some characteristics of Marfan syndrome such as overgrowth of long bones.

The tissues primarily affected by FBN1 mutation (including bone, aorta and pulmonary artery, mitral valve, zonular fibres of the eye, dura matter, skin and adipose) contain cells of mesenchymal origin, which synthesize connective tissue extracellular matrix (ECM), composed of fibrous proteins and glycosaminoglycans. The ECM provides strength and elasticity for these tissues. Fibrillin-1 is the major structural component of the extracellular microfibrils of the ECM10 and also seems to be involved in sequestering the growth factor TGFβ in inactive form.5,11,12 In adults, mesenchymal cells derive from stem cells residing in the bone marrow and mesenchymal tissues.13,14 These stem cells retain the ability to differentiate into cells of connective tissue lineages, including adipocytes, osteoblasts, chondrocytes, smooth and skeletal muscle, endothelial cells of blood vessels and fibroblasts (reviewed in Barry and Murphy14). Differentiation of mesenchymal cells into specific cell types requires induction of a range of transcription factors14 and may also involve interaction with cells of monocyte origin.15 During organogenesis, mesenchymal cells can also undergo transition to epithelial phenotype (mesenchymal–epithelial transition), with concomitant inhibition of mesenchyme-specific genes and activation of genes required to form intercellular adhesions characteristic of epithelium. The transition between the two states is regulated by a number of cellular factors, especially TGFβ family members.16,17 The actions of TGFβ on mesenchymal cells are mediated through transcription factors such as SNAIL and SLUG (encoded by SNAI1 and SNAI2 genes),18 and result in expression of mesenchymal genes and suppression of the epithelial marker E-cadherin.

The phenotype of Marfan syndrome is extremely variable, even among family members carrying the same mutation (see refs.4,19,20). Potential modifier genes for Marfan syndrome are likely to be found in the network of genes that are co-expressed in tissues affected by FBN1 mutation. Such genes would also be strong candidates for a role in diseases with related phenotypes. In this article, we identify and analyse genes that are stringently co-regulated with FBN1.

MATERIALS AND METHODS

Identification and annotation of an FBN1-associated cluster of genes

The analysis was performed on publicly available gene expression data (to which we contributed) generated from 44 mouse cell types and 2 mouse organs15,21 (Supplementary Table S1) using the Affymetrix MOE430_2
GeneChip and normalized using MASS (Affymetrix, Santa Clara, CA, USA). The data were accessed through GEO DataSets (accession number GSE10246).

Correlation networks were constructed from the data on the basis of pairwise Pearson’s correlation relationships. A network graph comprising 8578 nodes (probes) and 153 418 edges was generated using BioLayout Express3D. The resulting graph was then clustered using the Markov Clustering algorithm at an MCL inflation value of 1.7. Clustering was also performed on expression data from mouse tissues. Because the initial data did not include chondrocytes, which are likely to be involved in the skeletal phenotype of Marfan syndrome, the expression of cluster genes was also examined in data from a published study of chondrocyte differentiation (cultured limb bud mesenchymal cells; GEO Profiles accession no. GDS1863). In addition, we considered data sets from developing mouse kidney (E12.5; GEO Profiles accession no. GDS1583) and from developing mouse gastrointestinal tract (E18.5; GEO Profiles accession no. GDS2699).

**Location and function of cluster genes**

Genes in the cluster of interest were assessed for recognized homologies, cellular localization and function using publicly available databases (Ensembl, NCBI). Possible or verified involvement in disease was determined by searching the Online Mendelian Inheritance in Man (OMIM) and Online Mendelian Inheritance in Animals (OMIA) databases on the NCBI website.

**Determination of functional transcription factor binding sites in promoter regions of Fbn1-associated cluster genes**

The Affymetrix MOE430_2 probe set was mapped to mouse RefSeq genes and the beginning of RefSeq was taken as a predicted transcription start site. Bioinformatic analysis of motif activity and motif target predictions were performed as described previously. All genes represented on the microarray, which had been allocated a RefSeq (12752 in total), were classified as being either among the 205 genes of the Fbn1-associated cluster or not within the set. The proportion of genes with a z(p, m) score of greater than 1 for each transcription factor binding motif m was calculated for the two groups and a z-value for this difference was determined. This provides a measure of overrepresentation of predicted targets of the transcription factor in the mesenchymal cluster relative to other genes.

**RESULTS**

**Identification and annotation of Fbn1-associated genes in proliferating cells**

To identify genes that were strictly co-regulated with mouse Fbn1 in a cell-autonomous manner, we focused on a large data set derived from primary mouse cells, including primary calvarial osteoblasts undergoing differentiation and a range of haemopoietic cell types (see Supplementary Table S1), produced as described previously. Bioinformatic analysis of motif activity and motif target predictions were performed as described previously. All genes represented on the microarray, which had been allocated a RefSeq (12752 in total), were classified as being either among the 205 genes of the Fbn1-associated cluster or not within the set. The proportion of genes with a z(p, m) score of greater than 1 for each transcription factor binding motif m was calculated for the two groups and a z-value for this difference was determined. This provides a measure of overrepresentation of predicted targets of the transcription factor in the mesenchymal cluster relative to other genes.

**Fbn1-associated cluster genes in other data sets**

The cell lines assessed in the initial analysis did not include all mesenchymal cell types that would be found within tissues, nor all states of mesenchymal differentiation. To identify a subset of genes that were robustly expressed in mesenchymal tissues rather than cell lines, we clustered expression patterns across tissues in the publicly available GNF1M data set of gene expression in mouse tissues. Gene expression showed more diversity across these tissues and there was substantially greater noise in this data set, as evidenced by smaller clusters and overall lower correlation coefficients. Hence, a lower correlation level (r ≥ 0.75) was required to detect associations. A total of 119 genes clustered with Fbn1 in this analysis (Supplementary Table S3). Of these, 24 overlapped with the cluster derived from proliferating cells (indicated in Supplementary Table S3). Classic ECM genes such as Ebn, Fbn2, Mfap4, Mfap5 and Fbn2 also clustered with Fbn1 in this analysis of expression in tissues.

One major mesenchyme-derived cell type excluded from the cellular data was the chondrocyte. We therefore examined a published study of the differentiation of primary chondrocytes derived from embryonic footpads. Results for 235 of the 304 probes were available, representing 160 different genes. In all, 81% of these genes, including Fbn1, were in the highest 25% of expression at most or all time points, extending the view that these genes are co-expressed by proliferating mesenchyme, regardless of lineage.

As noted above, mesenchyme–epithelial transition is a key event in organogenesis. The transition has been analysed separately in developing mouse kidney and gastrointestinal tract. Fbn1 expression was strongly associated with mesenchyme in these data sets. Fbn1-associated cluster genes such as Bgn (biglycan), Cadl1 (caldesmon 1), Col1a2 (collagen type 1-2 subunit), Il6st (interleukin-6 signal transducer), Ror1 (receptor tyrosine kinase-like orphan receptor 1), Sparc (osteonectin; secreted protein, acidic, cysteine rich) and Timp2 (tissue inhibitor of metalloproteinases 2) showed a similar pattern of expression to Fbn1 in both the data sets, whereas others were profile neighbours of Fbn1 in one or the other data set (not shown).

**Cellular location and function of Fbn1-associated cluster genes**

As summarized in Table 1, 171 members of the Fbn1-associated cluster could be assigned a cellular location on the basis of experimental evidence or electronic annotation. The majority were extracellular but a substantial number were involved in secretion. For example, 10% of the annotated genes encoded proteins of the endoplasmic reticulum (including trafficking proteins and molecular chaperones), indicating a surprising level of target specificity for these processing proteins. Table 1 also shows that 181 genes could be assigned a function (Table 1). The largest group (25%) was a broad category of proteins involved in regulating cell size and number. A total of 10% were involved in ECM structure. There were 17 (9.4%) genes encoding known or putative transcription factors, including some families (SLUG/SNAI1, TWIST, PRRX, NFAT, ID, SOX) known to regulate mesenchyme differentiation or function. There were nine genes for expression and high variability (see expression profiles on BioGPS). Both mapped to intronic sequences (Affymetrix website) that have a low frequency of transcript initiation, indicating that these probes may detect rare variant Fbn1 transcripts that do not show clustering with the major Fbn1 probe sets. Probes for the other mouse fibrillin gene, Fbn2, which has overlapping functions with Fbn1, did not cluster with Fbn1 in this data set. Fbn2 showed expression only in osteoblasts and C3H 10T1/2 cells, and is therefore likely to function more specifically in bone.
Figure 1 Characteristics of the Fbn1-associated expression cluster. (a) A three-dimensional image of the Fbn1-associated cluster (nodes shown by black spheres, edges by grey lines) within the network. Other clusters are shown by edges only. (b) A two-dimensional image of the Fbn1-associated cluster with the two Fbn1 probe sets shown as black spheres. (c) Normalized expression of genes in representative cell types, averaged across all probes in the cluster. The means of two experiments performed in triplicate are shown. (d) Expression of two Fbn1 probes, 1460208_at (black) and 1425896_a_at (grey). The means of two experiments performed in triplicate are shown for each probe. Data are available at GEO DataSets (accession number GSE10246).
G-protein-coupled receptors and five for receptor tyrosine kinases. Most of the receptors had unknown ligands. Four genes had no informative annotation, with no similarity to known genes or assignable function or location.

The role of genes of the Fbn1-associated cluster in disease was examined by assessing entries in OMIM. Of 168 genes with an entry, 60 were associated with a phenotype in mouse (41) or human (29) (Supplementary Table S4). Of these, bones, skin, eyes and blood vessels were most frequently affected in humans, and bones, blood vessels and lung were most frequently noted for mouse. Eight mouse knockout models resulted in embryonic lethality. The results are consistent with a critical role for these genes in the development of the ECM.

**Determination of common functional transcription factor binding sites in promoter regions of Fbn1-associated cluster genes**

To assess the basis for their apparent co-regulation, we subjected the 205 genes of the Fbn1-associated cluster to an analysis of transcription factor binding sites.28 Table 2 lists the 15 transcription factors that had the highest positive correlations with the expression pattern of cluster genes. Comparison was also carried out between genes within the cluster and the remaining genes of the data set. Supplementary Table S5 shows the 65 transcription factor binding motifs that showed significant overrepresentation in the cluster genes. Seven transcription factor binding motifs showed a high correlation between activity and expression of the cluster genes and were consistently overrepresented in cluster gene promoters. The motifs were consensus sequences for binding proteins of the TEAD, RP58, MAZ, KLF4, IK1/IK2, BLIMP1 and CIZ families (Table 2). No Fbn1-associated cluster gene was significantly (Z > 2) associated with activity of all seven of these motifs, and Fbn1 alone was associated with six. Five of the genes were significantly associated with activity of five of these motifs and fifteen were associated with four.

**Identification of genes highly correlated with Fbn1**

When the initial clustering analysis using cell line data was repeated at a higher stringency of r ≥ 0.95, 46 probes (31 genes) were found to be in the same cluster as Fbn1 (Supplementary Table S6). Twelve of these genes were annotated as being located in the ECM, extracellular region or extracellular space. There were five recognized transcription factors and seven receptors. Eight of these genes had no or limited annotation, including a TGFβ-induced transcript (Tgfb1I1), a steroid-sensitive coiled-coil domain protein (Ccdk80) and a transmembrane protein (Tmem45a).

As noted above, in this study Fbn1 expression showed strong association (Z > 6.5) with the activity of six of the seven transcription factor motifs identified as having high activity in the cluster (for TEAD, RP58, MAZ, KLF4, BLIMP1 and CIZ family members; Table 2). Three genes (Lad3, Nfatc4 and Atoh8) were associated with activity of five of the six motifs in common with Fbn1 and 11 were associated with four of the six motifs in common with Fbn1 (Nua1k1, Col1a2, Col3a1, Gas1, Serpinh1, Cdhl1, Thbs2, Tpm1, Pcdh13, Boc, Grp23). In addition, Fbn1 expression was significantly associated (Z > 4) with a number of other motifs found to be overrepresented in the cluster. These included binding motifs for AP-4, MAZR, Broad Complex and SP1-gershenzon (Table 2). Five genes (Gas1, Capn6, Atoh8, Col1a2 and Sna12) were associated (Z > 2) with 10–15 of the same factors as Fbn1.

**DISCUSSION**

This analysis of gene expression data revealed that the mouse Fbn1 gene was in a cluster of 205 genes representing a lineage-independent expression signature for mesenchymal cells. Transcription factors binding TEAD, CIZ, RP58, MAZ, BLIMP1 and IK1/IK2 sites are candidate regulators of this Fbn1-associated cluster. Several of these have known roles in mesenchymal cell types. MAZ (myc-associated zinc-finger protein) has been shown to regulate muscle-specific gene expression.30 RP58 has an essential role in skeletal myogenesis,31 and BLIMP1 is involved in myocyte differentiation.32 CIZ is implicated in regulation of bone mass biology.33 TEAD2 and TEAD4, although not previously implicated in mesenchyme biology, had a similar expression pattern to the genes of the Fbn1 cluster. KLF family members and IK1/IK2 are associated with transcriptional repression in haemopoietic cells, and their function may be to prevent ectopic expression of the Fbn1-associated cluster genes in non-mesenchymal cells. Fbn1 was associated with six of these seven transcription factor motifs, the only gene with this level of association. The Fbn1-associated cluster itself includes genes for a number of transcriptional regulators that are known to be involved in epithelial-mesenchyme transition, including the Sna1, Sna2, Prx1, Prx2 and Twist1 genes. Our recent analysis34 detected motifs for PRRX family members in the Fbn1 proximal promoter region.

The study is limited by a number of factors. The published data were from cells of a single mouse strain, and it would be interesting to use a different strain, especially as there is considerable between-strain variability in gene expression (see Wells et al35 and mouse e-QTL data on BioGPS); we would predict that the same genes would continue to cluster on the basis of expression pattern, even though those patterns might vary with different strains. Many members of this Fbn1 cluster
(including BGN, key collagen genes, SERPINH1 and the transcription factor genes SNAI2, PRRX1 and TWIST1) were also co-expressed with FBN1 in human tumours and tissues (TC Freeman and TN Doig, unpublished results), as were several minimally annotated genes such as CD248, FKBP9, LRRC17 and TGFB1I1. We did not assess all cell types that are abnormal in Marfan syndrome. For example, the main morbidity comes from dissection of the aorta, and there were no aortic cells in the study, nor were there cells from the anterior segment of the eye or from dura mater. If these cell types were included, some of the genes would drop out of the cluster and those that remained would represent tightly co-regulated genes that are powerful candidates for a role in modulating the Marfan syndrome phenotype.

### Table 2: Transcription factor motifs showing the highest correlations of activity with expression of Fbn1-associated cluster genes and with Fbn1

<table>
<thead>
<tr>
<th>Transcription factor motif ID</th>
<th>Transcription factor family</th>
<th>Function</th>
<th>Correlation between motif activity and expression of cluster genes</th>
<th>Correlation between motif activity and Fbn1 expression</th>
</tr>
</thead>
<tbody>
<tr>
<td>MA0090</td>
<td>TEA domain (TEAD)/transcription enhancer factor (TEF)</td>
<td>Transcriptional activators, interact with vestigial homologues including Vgll342,43</td>
<td>20.01</td>
<td>10.53</td>
</tr>
<tr>
<td>M00532</td>
<td>58 kDa repressor protein (RP58)/zinc-finger protein 238 (ZNF238)</td>
<td>Neuronal transcription repressor</td>
<td>18.95</td>
<td>14.18</td>
</tr>
<tr>
<td>M00649</td>
<td>MYC-associated zinc-finger protein (MAZ)</td>
<td>Regulation of inflammation-responsive and muscle genes; regulates initiation and termination of transcription</td>
<td>16.96</td>
<td>11.57</td>
</tr>
<tr>
<td>M00483</td>
<td>ATF6</td>
<td></td>
<td>13.54</td>
<td>NS</td>
</tr>
<tr>
<td>MA0039</td>
<td>KLF4</td>
<td>Cell-cycle check-point regulator; transcriptional repressor</td>
<td>12.31</td>
<td>9.22</td>
</tr>
<tr>
<td>MA0083</td>
<td>SRF</td>
<td></td>
<td>11.24</td>
<td>NS</td>
</tr>
<tr>
<td>M00086</td>
<td>IK-1/IK-2/ZNFN1A</td>
<td>Chromatin remodelling</td>
<td>10.08</td>
<td>NS</td>
</tr>
<tr>
<td>M00087</td>
<td>Agamous</td>
<td></td>
<td>9.50</td>
<td>NS</td>
</tr>
<tr>
<td>M00243</td>
<td>EGR-1/NGFI-C/EGR-3/EGR-2</td>
<td></td>
<td>9.21</td>
<td>NS</td>
</tr>
<tr>
<td>M00244</td>
<td>Ciz2/ZNF384</td>
<td>Nucleocytoplasmic shuttling; suppresses bone formation</td>
<td>8.99</td>
<td>6.66</td>
</tr>
<tr>
<td>M00133</td>
<td>TST-1/POU3F</td>
<td></td>
<td>8.98</td>
<td>NS</td>
</tr>
<tr>
<td>MA0116</td>
<td>ROA2/ZNF423</td>
<td></td>
<td>8.81</td>
<td>NS</td>
</tr>
<tr>
<td>M00639</td>
<td>HNF-6/ONECUT2</td>
<td></td>
<td>8.27</td>
<td>NS</td>
</tr>
<tr>
<td>M01066</td>
<td>BLIMP1/PRDM1</td>
<td>Regulation of B and T cell and myocyte differentiation, transcriptional repressor</td>
<td>8.16</td>
<td>6.77</td>
</tr>
<tr>
<td>SPI1_gershenzon</td>
<td>SP1/Gershenzon</td>
<td></td>
<td>7.92</td>
<td>5.37</td>
</tr>
<tr>
<td>M00033</td>
<td>p300</td>
<td></td>
<td>7.45</td>
<td>5.61</td>
</tr>
<tr>
<td>M00005</td>
<td>AP-4</td>
<td></td>
<td>7.01</td>
<td>6.40</td>
</tr>
<tr>
<td>MA0113</td>
<td>Broad Complex 4</td>
<td></td>
<td>6.50</td>
<td>4.99</td>
</tr>
<tr>
<td>M00451</td>
<td>NKX3A</td>
<td></td>
<td>6.12</td>
<td>5.80</td>
</tr>
<tr>
<td>MA0029</td>
<td>EVI1</td>
<td></td>
<td>6.06</td>
<td>4.90</td>
</tr>
<tr>
<td>Homeobox_class:</td>
<td></td>
<td></td>
<td>5.77</td>
<td>4.30</td>
</tr>
<tr>
<td>Homeo MF0010</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MA0089</td>
<td>TCF11_MAF/C/NFE211</td>
<td></td>
<td>5.55</td>
<td>3.87</td>
</tr>
<tr>
<td>M00706</td>
<td>TFI-1</td>
<td></td>
<td>5.50</td>
<td>4.41</td>
</tr>
<tr>
<td>M00432</td>
<td>TTF1</td>
<td></td>
<td>4.64</td>
<td>4.29</td>
</tr>
<tr>
<td>M00491</td>
<td>MAZR</td>
<td></td>
<td>4.57</td>
<td>4.38</td>
</tr>
<tr>
<td>nanog_mm8</td>
<td>NANOG</td>
<td></td>
<td>4.45</td>
<td>4.26</td>
</tr>
<tr>
<td>octsox_dimer_mm8</td>
<td></td>
<td></td>
<td>3.65</td>
<td>2.59</td>
</tr>
<tr>
<td>sox2_human</td>
<td>SOX2</td>
<td></td>
<td>3.12</td>
<td>3.32</td>
</tr>
<tr>
<td>M00935</td>
<td>NF-AT</td>
<td></td>
<td>2.31</td>
<td>3.08</td>
</tr>
<tr>
<td>M00302</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00456</td>
<td>FAC1</td>
<td></td>
<td>2.20</td>
<td>2.27</td>
</tr>
<tr>
<td>M00418</td>
<td>TGIF</td>
<td></td>
<td>NS</td>
<td>2.20</td>
</tr>
</tbody>
</table>

NS, not significant (z<2).

*aTranscription factors or families known to bind to the consensus motif used in the search.

*bFunction given for the seven closely associated transcription factor families.

*c-z-value for the correlation between the activity profile for the motif and the average expression profile for the cluster.

*d-z-value for the correlation between the activity profile for the motif and the expression profile of Fbn1.

*European Journal of Human Genetics*
The rationale behind our analysis is that genes that are co-expressed with Fmn1 are candidate modifiers of the effects of FBN1 mutation in humans and may contribute to other diseases of connective tissue with similar phenotypes. In spite of the limitations, several examples validate this rationale. For instance, a strong association of Fmn1 with the Lox and Bgn genes was noted. Biglycan protein (encoded by Bgn) has been reported to stimulate synthesis of fibrillin-1 in pressure-induced renal injury and may have a more general role in assembly of connective tissue. No disease has been associated with BGN mutation in humans but Bgn-deficient mice have a skeletal phenotype. Lysyl oxidase (Lox) gene may be important in overall assembly of elastic microfibrils (reviewed in Wagenseil and Mecham) and may be involved in preparing tumour cells for metastasis. The human homologue of another of the cluster genes, SERPINF1, was recently implicated in a recessive form of osteogenesis imperfecta, a bone disease. Others within the cluster are not well characterized. They include novel transcription factors, G-protein-coupled receptors, nuclear receptors and receptor tyrosine kinases that are clearly potential drug targets and may be important in cell signalling during differentiation of mesenchymal cell types. A group of novel genes encoding hypothetical proteins was also present. The functions of these genes can now be inferred from their co-expression in the cluster and clearly warrant a detailed characterization in cells of mesenchymal lineage, and also consideration as modifiers of the Marfan phenotype or as candidate genes in human connective tissue diseases.

CONFLICT OF INTEREST
The authors declare no conflict of interest.

ACKNOWLEDGEMENTS
This work was supported by a grant-in-aid from the National Heart Foundation of Australia (G06R8527) and by the Biotecnology and Biological Sciences Council of the United Kingdom (Institutional Strategic Programme Grant to The Roslin Institute).

WEB RESOURCES
The URLs for data presented herein are as follows:


Supplementary Information accompanies the paper on European Journal of Human Genetics website (http://www.nature.com/ejhg)