From SNP to Transcriptional Mechanism: A Model for FRMD3 in Diabetic Nephropathy

Sebastian Martini¹, Viji Nair¹, Sanjeevkumar R. Patel¹, Felix Eichinger¹, Robert G. Nelson², E. Jennifer Weil², Marcus G. Pezzolesi³, Andrzej S. Krolewski³, Ann Randolph¹, Benjamin J. Keller⁴, Thomas Werner¹,⁵, Matthias Kretzler¹

¹Departments of Internal Medicine – Nephrology, University of Michigan, Ann Arbor, Michigan; ²National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Phoenix, Arizona; ³Research and Clinic Divisions, Joslin Diabetes Center, Boston, Massachusetts; ⁴Department of Computer Science, Eastern Michigan University, Ypsilanti, Michigan; ⁵Genomatix Software GmbH, Munich, Germany

Corresponding author:
Matthias Kretzler
University of Michigan
Dept. of Internal Medicine
1150 W. Medical Center Drive
1560 MSRB II
Ann Arbor, MI 48109-0676
Tel.: (734) 615-5757
Fax: (734) 763-0982
E-MAIL: kretzler@umich.edu

Word count: 4509
Figures: 4 Suppl. Tables: 2

Short title: FRMD3-SNP impact on DN
Abstract

Genome wide association studies (GWAS) have proven to be highly effective at defining causal relationships between single nucleotide polymorphisms (SNP) and clinical phenotypes in complex diseases. Establishing a mechanistic link between a non-coding SNP and the clinical outcome is a significant hurdle in translating associations into biological insight. We demonstrate an approach to assess the functional context of a diabetic nephropathy (DN) associated SNP located in the promoter region of the gene FRMD3. The approach integrates pathway analyses with transcriptional regulatory pattern based promoter modeling and allows the identification of a transcriptional framework impacted by the DN-associated SNP in the FRMD3 promoter. This framework provides a testable hypothesis for mechanisms of genomic variation and transcriptional regulation in the context of DN. Our model proposes a possible transcriptional link through which the polymorphism in the FRMD3 promoter could influence transcriptional regulation within the bone morphogenetic protein (BMP) signaling pathway. These findings provide the rationale to interrogate the biological link between FRMD3 and the BMP pathway, and serve as an example of functional genomics-based hypothesis generation.

(Words: 174)
**Introduction**

While genome wide association studies are effective at projecting genetic variants to complex disease phenotype, establishing the corresponding mechanistic link remains difficult. This is especially true for SNPs in non-protein coding regions of the genome that may impact regulatory function in a manner that is only evident in a particular functional context (1). One such context may be a biological process determined by genes whose transcription is synchronized by common regulatory elements within their promoters (2; 3). A SNP located in one of these regulatory elements may alter or disrupt this coordinated regulation, leading to a change in gene expression and subsequently phenotype. It may be possible to identify such a mechanism via a change to a transcription factor binding site (TFBS) by a candidate SNP; we demonstrate this strategy for a SNP affecting the diabetic nephropathy-associated bone morphogenetic protein (BMP) signaling pathway. The approach allows us to generate testable hypotheses from GWAS candidates falling in promoter regions, and has the potential to help understand the functional impact of genetic variants in diabetic nephropathy and other complex genetic diseases.

Diabetic nephropathy (DN) is a complex disease and the leading cause of end-stage renal disease in the U.S. (4), and about 20-40% of all patients with either type 1 (T1D) or type 2 diabetes (T2D) develop DN (5-7). DN has a significant heritability (8), providing the rationale for performing GWAS to discover genetic loci implicated in DN (9). Initial DN GWAS discovered candidate genetic loci for predisposition to DN for both T1D and T2D (8; 10). However, these associations of a locus with DN do not explain how associated alleles affect the mechanism of disease. Unfortunately, this situation is typical of most
GWAS of complex genetic disorders, while loci whose effects have been functionally confirmed are generally associated with Mendelian disorders. An example is the autosomal dominant disorder multiple osteochondromas, for which a SNP located in the EXT1 promoter, eliminates a TFBS and increases promoter activity (11). For complex diseases, any large-scale analysis involving luciferase assays, electrophoretic mobility shift assays (EMSA) and enzyme-linked immunosorbent assays (ELISA) are simply not feasible for hundreds of diseases-associated SNPs. Data-driven approaches including the one outlined in this manuscript are necessary to prioritize the number of testable hypotheses for further experimental validation.

Establishing the functional context of a SNP is important in defining such hypotheses. Our group has previously employed a functional context approach to identify proteins associated with the glomerular slit diaphragm in DN (12). In that work, a regulatory module detected in the promoters of a few known slit diaphragm genes predicted other slit diaphragm molecules after a genome-wide promoter search. Here, our integrative approach combines regulatory SNP prediction, transcriptional promoter modeling and pathway analysis capable of decoding putative transcriptional pathomechanisms of DN (Fig. 1). We focus on the candidate gene FERM domain containing 3 (FRMD3) identified by a GWAS of the Genetics of Kidneys in Diabetes (GoKinD) study collection (13). In that study, the SNP rs1888747 showed the strongest risk association ($P=4.7\times10^{-7}$, OR=1.45) with DN within T1D subjects. Despite different study designs, this SNP also reached statistical significance level in a replication study of 1,305 participants of the DCCT/EDIC study, as well as in a subcohort of Japanese subjects with T2D (14). This
polymorphism remained significantly associated with DN in a random-effects meta-analysis of genetic variants reproducibly associated with DN (15). Additionally, we have recently shown that rs1888747 is significantly associated with DN among 66 large T2D families from the Joslin T2D family collection (Pezzolesi et al., in preparation). The SNP rs1888747 is located on chromosome 9q in the extended promoter region of FRMD3. FRMD3 has not previously been implicated in the pathogenesis of DN, T1D or T2D.

Here we describe both our in-silico approach, and its use to derive the hypothesis that a DN risk allele brings FRMD3 under the control of a proposed transcriptional regulatory module and inhibits renal expression of FRMD3. Not only does the approach detect a transcriptional regulatory pattern affected by the candidate SNP, it also connects known DN-associated pathways to the GWAS-derived candidate gene, providing the testable model system for further insight into the pathophysiology of DN.

**Research Design and Methods**

**Strategy**

We hypothesized that the SNP rs1888747, reported to be associated with DN by Pezzolesi et al. (13) and located in the proximity of FRMD3, is a regulatory SNP that alters the transcription factor binding capabilities of the FRMD3 proximal promoter region. We assumed that the binding site putatively affected by the SNP is part of a molecular TFBS framework involved in this transcriptional change, which should also be conserved in promoters of functionally connected (i.e. co-varying) transcripts. Finding those transcripts might enable us to detect the framework by including the polymorphism in the
FRMD3 promoter, and thus place the SNP into a DN-relevant functional context.

We used comparative promoter analysis to determine common regulatory elements of FRMD3 and its co-expressed transcripts. Promoters of functionally linked transcripts are likely to contain conserved (non-random associated) TFBS frameworks. SNP-related TFBS alterations have the potential to integrate genomic features with transcriptional regulatory functions. A detailed overview of our study and strategy can be found in Fig. 1.

**Human Renal Biopsies**

Renal biopsy samples were procured from 22 participants in a clinical trial, whose extended follow-up provides an opportunity to examine the etiology of DN in T2D as well as the effect of treatment with losartan on the onset and progression of diabetic kidney disease (Study Title: Renoprotection in Early Diabetic Nephropathy in Pima Indians, ClinicalTrials.gov Identifier: NCT00340678). Renal Biopsy specimens were processed and analyzed as described previously (12; 16; 17). Subjects' aggregate clinical and histological characteristics are summarized in Suppl. Table 2.

**FRMD3 expression in subjects with DN and either normal or decreased GFR**

We compared glomerular FRMD3 expression levels as well as individual estimated glomerular filtration rate (eGFR) measurements of 22 Pima Indians with normal GFR to a cohort of 7 T2D subjects with CKD stage 3 to assess whether FRMD3 gene expression would correlate with renal function. Statistical analysis comparing the two groups was
done using GraphPad Prism 5 with a two-tailed t-test (Mann-Whitney U, 95% confidence interval). A P<0.05 was considered statistically significant.

**Pathway analysis of FRMD3 co-expressed transcripts**

When genes are co-regulated under various biological conditions, their corresponding expression profiles may show relative similarity, or co-expression (18). We identified FRMD3 co-expressed transcripts by calculating the Pearson’s r correlation between the expression profiles of FRMD3 and all other genes expressed above background. These co-expressed, potentially co-regulated transcripts were then analyzed to identify transcripts known to be functionally related using Ingenuity Pathway Analysis software (IPA, version 8.5, Ingenuity® Systems, Redwood City, CA, http://www.ingenuity.com). The software detects enriched canonical pathways in a given gene set. Default settings were applied.

**Renal function associated with FRMD3 co-expressed transcripts**

An unsupervised hierarchical clustering analysis of the 22 Pima Indians (T2D DN) using the expression levels of 581 FRMD3 co-expressed genes (including FRMD3) was performed (MeV version 4.5.1, Euclidean distance, average linkage method). The two main branches in the dendrogram showed 100% support (bootstrap, n=1000). They were further analyzed for differences in their FRMD3 expression and their ability to associate with clinical and histologic subgroups as this would link FRMD3 co-expressed transcripts with a disease-associated phenotype. Renal function measures, iGFR (mL/min) measured by a urinary clearance method that used cold iothalamate (19), the albumin creatinine
ratio (ACR) and the fractional mesangial area were compared between the two clusters. The delta in ACR/year and the delta in iGFR (mL/min)/year were calculated by subtracting the corresponding value from the time of enrollment into the study from the latest available value divided by the number of years of follow up. Fractional mesangial area was determined as described previously (20). Statistical analysis comparing the two major cluster branches was done using GraphPad Prism 5 with a two-tailed t-test (Mann-Whitney U, 95% confidence interval). A P<0.05 was considered statistically significant.

**Computational Promoter Analysis and Evaluation**

Promoter regions for the eight FRMD3 co-expressed BMP pathway members were extracted (Genomatix® ElDorado database version 07/2009) and promoter modeling was performed to detect common transcriptional regulatory elements potentially influenced by the SNP of interest. For the FRMD3 promoter, we extracted a sequence of +/- 320 nucleotides (nt) around the SNP of interest, rs1888747. A sequence length of 320 nt was chosen to allow the detection of a four element promoter module starting at the SNP position with an estimated average distance of 80 nt between the centers of two consecutive elements. The SNP rs1888747 is located at position 85345371 on chromosome 9 (Genome Build 36.3) in the extended promoter sequence of FRMD3 (1904 nt proximal to the first transcription start region). We determined potential TFBS generated or lost by the SNP rs1888747 (Genomatix® MatInspector) as described by Cartharius et al. (21). The FRMD3 promoter sequence was analyzed both with and without the risk allele. A promoter module is defined as a set of two or more TFBS of a defined order, orientation and distance range acting together in a certain functional
context (see *Fessele et al.* (2)).

We searched for a common module among promoter sequences of a subset of the 8 *FRMD3* co-expressed *BMP* pathway members and the SNP-altered sequence of the *FRMD3* promoter (Genomatix® FrameWorker). Variance, and distance between the individual promoter elements were altered until a module with more than two elements was discovered. We required more than two elements to be identified in our search since more complex modules have been shown to be associated with more specific biological function (22). In addition, the promoter module was required to occur in at least 2 of the 8 *FRMD3* co-expressed *BMP* pathway members as well as in the *FRMD3* promoter sequence at the position of rs1888747.

We evaluated the significance of the promoter module by searching a genome wide human promoter database for additional genes whose promoters would also contain potential binding capabilities for the defined framework identified in the previous step (Genomatix® ModellInspector). To achieve comparable preconditions, this search was conducted after adjusting the promoter sequence of all genes in the promoter database (Genomatix® ElDorado Version: 7-2009, 93,372 promoters) to the same sequence length where rs1888747 was found in the promoter of *FRMD3*. Additional *BMP* pathway members identified by this approach were evaluated for their enrichment in comparison to the total number of additional detected genes.

**EMSA**
EMSA was conducted to evaluate protein-binding differences of the FRMD3 wild type and SNP-altered sequence. While this method does not allow conclusions about the actual binding protein itself, it is an effective way for an initial assessment of regulatory capabilities of a SNP in a non-coding region. The following steps were taken:

a) *Glomerular Isolation.* Glomeruli from five three-month-old C57BL/6J mouse kidneys were isolated (23) with modifications in the nylon membranes used (100-μm nylon sieve, Sefar, Briarcliff Manor, NY).

b) *Nuclear extracts.* Nuclear protein extracts from adult mouse kidneys and livers, glomeruli isolated from adult murine kidneys, and 293 cells were prepared as described previously (24).

c) *EMSA Analysis.* Oligonucleotides corresponding to the wild type DNA sequence 5'-ACAAGGCTCTGGGAAACCAA C TGGCCATTGTCAACAATA43', or to the SNP sequence 5'-ACAAGGCTCTGGGAAACCAA G TGGCCATTGTCAACAATA43' and complimentary strands, were annealed and end-labeled with 32-P-dCTP (24). Nuclear protein extracts were incubated in buffer with poly dIdC or poly dAdT and 10,000 cpm of end-labeled oligonucleotide as described previously (24). For competition experiments, unlabeled DNA was added to the binding reactions at a 100-fold excess of the radiolabelled oligonucleotide. The DNA–protein complexes were resolved on 6% non-denaturing polyacrylamide gels in Tris–Borate–EDTA buffer at 120V for 2.5 hours. Gels were dried and exposed to XOMAT film (Eastman Kodak) overnight. The intensity of the DNA-protein complex was measured using the software ImageJ 1.44p from the NIH (http://imagej.nih.gov/ij/). A paired t-test (GraphPad Prism 5) was used to assess the
significance of the mean intensity in the SNP sequences compared to the WT-sequence.

Results

Defining clinical and functional association of FRMD3

To assess the functional relationship between FRMD3 and DN, we related steady state mRNA levels with the available clinical outcome parameters. We found FRMD3 transcript levels decreased significantly with progression of DN (8.9±1.2 in DN with CKD III compared to 10.3±0.9 (mean±standard deviation) in DN with normal GFR (P<0.02)) (Fig. 2 A).

As FRMD3 had no prior link to DN, we used a data driven approach to establish a putative clinical and functional context for FRMD3 in DN. Starting from a list of 17,589 transcripts expressed on the Affymetrix microarray chip, 16,956 passed the cutoff filter (median+2*SD of the 27 Poly-A Affymetrix negative controls’ expression baseline (25)) and were tested for correlation with FRMD3. Transcriptional co-regulation orchestrated by common upstream transcriptional regulatory elements (2) provided the rationale that FRMD3-correlated transcripts (similar mRNA expression patterns) might be linked to regulatory pathways in DN, which in turn may help establish the link between FRMD3 and the disease.

We identified 581 FRMD3 co-expressed transcripts (|r|≥0.65, FDR≤0.02; for top 10 transcripts with the highest |r|-value, see Suppl. Table 1). The majority (518) of the 581 FRMD3 co-expressed transcripts were concordantly regulated with FRMD3, as were the top 10 (sorted by their |r|-value) FRMD3 co-expressed transcripts. For 5 of those top 10
transcripts or close variants, an association with diabetes, cardiovascular or inflammatory diseases has been published (Suppl. Table 1), consistent with the relevance of this gene set to the pathophysiology of DN.

**Expression of FRMD3 and its correlated transcripts is linked to early progression in DN**

Hierarchical clustering using the expression signatures of FRMD3 co-expressed transcripts detected 2 distinct clusters (Fig. 2 B, C). Patients contained in the cluster 1 had a significantly (P=0.017) higher delta ACR/year of 212.4±227.9 (mean±standard deviation) compared to cluster 2 (delta ACR/year of 3.7±8.7, Fig. 2 D). Mesangial expansion, a key histologic feature of DN (20) was significantly (P=0.04) increased in cluster 1 (30±14%) compared to cluster 2 (17±7%, Fig. 2 E). Delta GFR showed a similar trend, but missed statistical significance. Observation times were similar in both patient groups (cluster 1: 9.0±2.2 years, cluster 2: 9.5±0.9 years, P=0.91). In cluster 1 with higher delta ACR/year, the gene expression of 7 out of the 8 BMP pathway genes (BMPR2, CREB1, KRAS, MAP3K7, PRKAR2B, SMAD5, and XIAP) was lower than in cluster 2. This concordance of transcriptional regulation of FRMD3 and BMP pathway members with renal outcome measures points towards a common molecular mechanism responsible for the co-regulation of FRMD3 and several BMP pathway members.

**Pathway analysis of FRMD3 co-expressed transcripts**

We determined the functional context of FRMD3 and its 581 co-expressed transcripts by mapping them to known canonical pathways. Among them, the BMP signaling pathway
was found to be the pathway with the strongest enrichment with eight BMP pathway members co-expressed with FRMD3 (BMPR2, CREB1, KRAS, MAP3K7, PRKAR1B, PRKAR2B, SMAD5, and XIAP) (Fig. 3). This finding is consistent with previous publications attributing DN-protective properties to the BMP pathway (for review, see (26; 27)) and indicates that the biological context defined for FRMD3 and its co-expressed transcripts might indeed be relevant for DN.

**Defining putative SNP function**

*In-silico* comparison of sequence variants with and without the risk allele identified a potential homeodomain factor (HOMF) TFBS covering the SNP position. This TFBS was not detected in the presence of the non-risk allele in the FRMD3 promoter (Fig. 1, step 2). An EMSA of oligonucleotides corresponding to the wild type and SNP-altered sequence of glomerular extracts from C57Black6 mice supports these predictions: the sequence with the disease-associated SNP shows a more than 4.74 times relative increase (intensity wild type (WT) vs. SNP: 1 vs. 57, 15 vs. 92, 31 vs. 145) in protein binding compared to the WT-DNA sequence (Fig. 4). These results show that rs1888747 affects protein binding, suggesting the generation of a putative TFBS by that particular SNP.

**Putative transcriptional mechanism for co-regulation of FRMD3 and BMP pathway members**

After extraction of the proximal promoter sequences of the eight BMP genes co-expressed with FRMD3, we identified promoter frameworks shared among BMP genes as well as the FRMD3 promoter sequence with the risk allele. For FRMD3 and four of the
eight FRMD3-co-expressed BMP pathway members (XIAP, KRAS, PRKAR2B, MAP3K7) we found a module with four TFBS (HOMF, BRNF, BRN5, GATA) where the SNP rs1888747 occurs in the first (HOMF) TFBS of FRMD3 (for details of the framework see Fig. 1, step 6). This framework provides the molecular basis for a proposed co-regulatory pattern of FRMD3 and BMP pathway members. A genome-wide search in a human promoter database (Genomatix® ModellInspector/ElDorado) identified an additional set of 18 BMP pathway members containing the 4 TFBS-module in their promoters. An enrichment analysis showed that detecting the promoter module in 22 (18 newly identified plus 4 original BMP pathway members) out of the total 72 BMP pathway genes as annotated by Ingenuity (Ingenuity Pathway Analysis software) achieved an enrichment score of 4.2 and a significant z-score of 7.6. These findings suggest that the 4 TFBS promoter module could mediate the transcriptional co-regulation of BMP pathway members and FRMD3 in the functional context of DN. Our results provide a rationale and an experimental framework to define a regulatory link between FRMD3 and the BMP pathway in DN.

Discussion

With the emerging capabilities to capture the genetic and molecular underpinnings of diabetic complications, molecular based disease definition can lead to individual risk assessments and selection of targeted therapies (28). Describing gene-environment interactions will be a critical step towards molecular disease definition. A series of studies currently aims to link genetic variation to diabetic complications (13; 29-31). Genetic variants can impact the phenotype by directly altering the coding sequence of a gene,
resulting in a qualitative change in the encoded protein. Alternatively, variants can alter regulatory regions in the genome resulting in quantitative changes of the transcript. Research in monogenetic diseases has established a clear path forward to define the consequences of protein coding variants. Defining the consequences of regulatory variants on gene expression, particularly in complex diseases, is still in its infancy. The present study aims to provide one possible way forward to identify potential regulatory effects of DN associated non-coding variants and their link to complex regulatory networks in DN.

Regulatory network analysis starting from a putative causal SNP needs to be embedded in an in-depth analysis of the functional context of the affected gene. This context is required to reveal regulatory mechanism represented by transcription factor binding site frameworks active in regulatory regions of the genes of interest. In general, regulatory SNPs can be inferred, if a known or potential TFBS is directly affected by the polymorphism (32). However, since individual TFBS are often not sufficient for regulatory functions, their functional contributions can only be assessed in the appropriate regulatory context, i.e. the interaction with other TFBS (33). Disease-relevant pathways and transcriptional co-variance can serve as selection criteria for genes belonging to that functional context. Regulatory links identified by this approach allow prediction of transcriptional alterations, which can be tested in the context of disease.

This strategy presented in our study is applicable whenever a transcriptional change of the GWAS gene is observed and co-regulated transcriptional networks can be identified. However, although this implies finding a group of co-expressed genes, the pathway association might not always be as clear-cut as in our case, which might result in testing
multiple associated pathways with the strategy presented above. A direct hit of the SNP in a TFBS is an advantage but proximity to a potential TFBS framework most likely would suffice to alter TFBS function. In case no such framework can be found with any associated pathway, alternative bioinformatics methods for the selection of genes of a similar functional context can be tested: including protein-protein interaction networks (34), phylogenetic conservation (12) or epigenetic/epigenomic approaches (35). With the increasing availability of genetic mapping of expression quantitative trait loci (eQTL) studies in DN cohorts, eQTLs will be linked directly to the physical location of transcripts differentially expressed in DN and thereby support promoter modeling approaches as described by our example (36; 37).

The study presented here started from a worst case scenario, as a testable hypothesis had to be developed for the role of a non-coding SNP in a gene without known function in DN. We followed a sequential strategy integrating multiple lines of genetic and genomic evidence for hypothesis generation (see Fig. 1 for overview). First, the candidate SNP rs1888747 in the proximal promoter region of FRMD3 prompted us to search for the functional context of the TFBS framework covering the candidate SNP. Pathway analysis of co-expressed transcripts revealed a significant enrichment for the BMP pathway (38). BMPs are part of the transforming growth factor β superfamily (39) and have a well-established role in kidney development, cell growth, cell differentiation, chemotaxis and apoptosis of various cell types (40). An imbalance of BMP-7 agonists like Kielin/chordin-like protein and BMP-7 antagonists like Gremlin has been described in DN (27). Decreased expression of BMP-7 and its agonists has been associated with increased
profibrotic activity in animal models of DN (41), consistent with a protective effect of BMP activation in DN. Promoter modeling for the FRMD3 promoter sequence as well as for 8 co-expressed transcripts led to the discovery of a BMP pathway-specific TFBS framework that identified a total of 22 BMP pathway members in a genome-wide promoter sequence search.

Our results support the hypothesis of a functional connection of the SNP with reduced FRMD3 expression, as the SNP-created binding site is located in a likely repressive promoter module. Since this module is shared between regulatory regions of 22 genes of the BMP pathway, BMP genes could be suppressed by the same mechanism utilizing the shared module. The risk allele generates the necessary binding sites of the BMP module in the FRMD3 promoter and, as for BMPs, represses FRMD3 with deleterious impact on DN including inhibition of the protective effects of the BMP pathway. Interestingly, a BMP-focused candidate gene study by the UK Genetics of Kidneys in Diabetes (GoKinD) Study Group was not able to identify statistically significant DN-associated SNPs in the genes BMP2, BMP4 and BMP7 (42). The above hypothesis establishes a trans-association of the DN-associated SNP linking BMP genes to the risk of DN via FRMD3.

**Proposed model connecting FRMD3 and BMP pathway**

Based on our findings, we developed a testable hypothesis for the functional impact of the SNP rs1888747 in DN. We suggest that our proposed TFBS framework is generally inhibitory in the context of renal gene expression and may act as a negative regulatory
feedback-loop to balance BMP pathway action. A maximum parsimony of all known facts is consistent with the idea that one FRMD3 function is to aid in the activation of BMP pathway gene expression, providing some counter-balance to the inhibitory effect of the TFBS framework defined for BMPs above. This is consistent with the observed higher expression of BMP genes in the absence of the risk allele. However, the risk allele brings FRMD3 under the control of the same negative BMP feedback loop effectively abolishing the positive impact of FRMD3 on BMP expression. As a result, BMP mediated protective effects on renal tissue, and thus renal protection, are reduced in individuals with the polymorphism, which is consistent with the observed DN phenotype associated with the polymorphism. FRMD3 and BMP pathway gene repression is correlated to the severity of the renal phenotype. Recent GWAS of T2D subjects also detected SNPs in the FRMD3 gene region to be associated with diabetic retinopathy, possibly relating to a uniform connection of FRMD3 and BMP pathway members in diabetes end organ damage (43).

The strength of this approach is its ability to predict functional connections based solely on regulatory networks as exemplified by significantly enriched transcriptional TFBSs frameworks in the absence of direct protein-based evidence. We currently do not know how the connection between FRMD3 and BMP pathway members is mediated. We found no evidence at the protein-, RNA- or miRNA-level. Therefore, FRMD3 is thought to influence currently unknown regulatory intermediates. Even in this case, the model provides an explanation of how this SNP could bring the transcriptional regulation of FRMD3 under the same control as the co-regulated BMP genes via the 4 TFBS regulatory
module. While beyond the scope of our manuscript, functionality can now be established experimentally in vivo. The model approach introduced here provides insight into genomic variation and the mechanisms of transcriptional regulation and provides the basis for targeted experimental design. FRMD3 appears to be a promising target for these experiments as comparative genome mapping data also confirmed FRMD3 as a nephropathy candidate gene in mice (44). The functional context proposed in this study could be experimentally validated by several approaches. Luciferase promoter reporter assays corresponding to wild-type and disease-associated alleles could be used to determine the functional impact of the rs1888747 SNP on FRMD3 expression and functional consequences of FRMD3 gene silencing/overexpression on the expression of BMP pathway members can be tested in vitro. The impact of the polymorphism in DN in vivo can be evaluated using mice transgenic for the FRMD3 locus with and without the disease-associated polymorphism. As our data provide a functional link of BMP signaling pathway members to other potentially DN-associated pathways such as the IGF-1 and Insulin receptor signaling pathway, results from these functional assays can be interpreted with regard to all pathways shown to be enriched among FRMD3-correlated transcripts.

**Conclusion**

Our work provides a paradigm of how functional genomics based hypothesis generation can be implemented by a stepwise integration of regulatory SNP prediction, transcriptional promoter modeling and pathway analysis. Our model approach provides a novel strategy to extend insight into the mechanisms of genomic variation and transcriptional regulation to regulatory networks informing subsequent experimental
design. The general approach can be applied for different questions in the field of GWAS and transcriptomic data integration. The method is also suitable for the analysis of experimentally derived TFBS data sets, such as ChIP-Seq data or panels of \textit{in vivo} protein bound DNA elements, generated by genomic footprinting (45). Furthermore, information from chromatin histone modifications, potentially regulatory sequences, or phylogenetic footprinting studies can be linked to regulatory networks. In the context of DN our work presents a novel starting point for hypothesis generation in molecular medicine in DN.

\textbf{Acknowledgments}

This study was supported in part by National Institutes of Health grants R01-DK058549 to A.S.K., K01-DK090125 to M.G.P., by the O’Brien Renal Center (P30-DK081943), and the National Center for Integrative Bioinformatics (U54-DA021519) to M.K. and by grant 01EX1021L to T.W. S.M. is the guarantor of this work and, as such, had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis. Parts of this study were presented at the 2010 meeting of the International Society of Nephrology in Geneva, Switzerland. We thank Claudiu V. Komorowsky, University of Michigan, Internal Medicine, Nephrology for helpful discussion of the renal morphometric data, as well as our collaborators from University College Dublin (Eoin P. Brennan, Catherine Godson, Finian Martin) for discussion on functional relationships between FRMD3 and BMP pathway members. No potential conflicts of interest relevant to this article were reported.
Author Contributions

The project was conceived and the experiments were designed by S.M, V.N, and M.K. Transcriptomic data analysis was performed by S.M, V.N., F.E. and A.R. The EMSA assays were conducted by S.P.. R.G.N. and E.J.W. phenotyped the Pima Indian participants, provided data and tissue samples for gene expression studies. E.J.W. performed the morphometric characterization of the kidney tissue. The manuscript was prepared by S.M., T.W., V.N, B.J.K, M.G.P., and S.P.. M.K., E.J.W. and R.G.N. reviewed and edited the manuscript. The study was supervised by M.K. and A.S.K.

References


Figure Legends

**Fig. 1:** Overview of the analysis strategy (2-7) to identify the putative regulatory effect of GWAS candidate (1) on FRMD3 regulation (2), linking the gene to transcriptional regulation of the BMP pathway (4-7) and DN, and suggesting a hypothetical regulatory model (8).

**Fig. 2 A:** *FRMD3* is repressed with progression of DN.
*FRMD3* gene expression comparing 22 Pima Indians with T2D and normal GFR with a cohort of 7 T2D with CKD stage 3. Data are displayed as mean±standard deviation. A) Glomerular *FRMD3* expression in early DN (Pima): 10.3±0.9, in CKD3 DN: 8.9±1.2. eGFR in early DN: 104±19 mL/min/1.73m², CKD stage 3 DN subjects : 53±33 mL/min/1.73m² (P<0.002, Mann-Whitney U, 95% confidence interval).

**Fig. 2 B:** *FRMD3* coregulated genes segregate DN patients in defined subgroups.
Cluster dendrogram of 581 *FRMD3*-correlated genes (including *FRMD3*) in a cohort of 22 Pima Indians with T2D diabetic nephropathy. The two main branches (cluster 1 and cluster 2) of the dendrogram show 100% support and reflect distinct clinical groups (see Fig. 2D).
Fig. 2 C: *FRMD3* and co-regulated *BMP* pathway members are repressed in cluster 1.  
*FRMD3* and *BMPR2, CREB1, KRAS, MAP3K7, PRKAR2B, SMAD5, XIAP* (7 of 8 *BMP* pathway members) are significantly (**P<0.008**) down-regulated in cluster 1 compared to cluster 2 (Mann-Whitney U, two tailed, 95% confidence interval). Expression data are displayed as mean±standard deviation. Glomerular *FRMD3* expression cluster 1: 8.29±0.54, cluster 2: 9.67±0.41.

Fig. 2 D: *FRMD3/BMP* repression is associated with increase of albuminuria.  
Clinical measures of delta ACR/year comparing the 2 main cluster branches from Fig. 2B. Data are displayed as mean±standard deviation. Delta ACR/year cluster 1: 212.4±227.9 is significantly (**P=0.017**) increased compared to Delta ACR/year in cluster 2: 3.7±8.7. (Mann-Whitney U, two tailed, 95% confidence interval).

Fig. 2 E: *FRMD3/BMP* repression is associated with increase of fractional mesangial area.  
Histologic measures of fractional mesangial area (%) comparing the 2 main cluster branches from Fig. 2B. Data are displayed as mean±standard deviation. Mesangial expansion was significantly (**P=0.04**) increased in cluster 1 (30±14%) compared to cluster 2 (17±7%) (Mann-Whitney U, two tailed, 95% confidence interval).

Fig. 3: Functional association of *FRMD3*-correlated genes.  
Top 10 pathways (Ingenuity Pathways Analysis, Ingenuity® Systems) of 581 *FRMD3*-correlated genes sorted by the ratio of members of the pathway among *FRMD3*- correlated genes vs. total number of members of that pathway. P≤0.001 for all pathways.

Fig. 4: Increased binding of glomerular nuclear extracts to DN-associated genomic region.  
Electrophoretic mobility shift assay (EMSA) from oligonucleotides corresponding to the wild type DNA sequence (WT) and SNP-altered sequence (SNP) of glomerular extracts from C57Black6 mice. The nonspecific competitor poly dIdC was used. Arrow indicates position of protein-bound oligos. With increasing amounts of protein used a distinct binding signal can be detected in the SNP-sequence but to a lesser amount in the WT-sequence as displayed in the Intensity Blot B). Intensity of the DNA-protein complex in lane 2 was set to 1.0. A paired t-test showed that the mean intensity was significantly higher in the SNP sequences compared to the WT-sequence (P=0.04).
Fig. 1: Overview of the analysis strategy (2-7) to identify the putative regulatory effect of GWAS candidate (1) on FRMD3 regulation (2), linking the gene to transcriptional regulation of the BMP pathway (4-7) and DN, and suggesting a hypothetical regulatory model (8).

180x282mm (300 x 300 DPI)
Fig. 2 A: FRMD3 is repressed with progression of DN.
FRMD3 gene expression comparing 22 Pima Indians with T2D and normal GFR with a cohort of 7 T2D with CKD stage 3. Data are displayed as mean±standard deviation. A) Glomerular FRMD3 expression in early DN (Pima): 10.3±0.9, in CKD3 DN: 8.9±1.2. eGFR in early DN: 104±19 mL/min/1.73m2, CKD stage 3 DN subjects: 53±33 mL/min/1.73m2 (P<0.002, Mann-Whitney U, 95% confidence interval).

Fig. 2 B: FRMD3 coregulated genes segregate DN patients in defined subgroups.
Cluster dendrogram of 581 FRMD3-correlated genes (including FRMD3) in a cohort of 22 Pima Indians with T2D diabetic nephropathy. The two main branches (cluster 1 and cluster 2) of the dendrogram show 100% support and reflect distinct clinical groups (see Fig. 2D).

Fig. 2 C: FRMD3 and co-regulated BMP pathway members are repressed in cluster 1.
FRMD3 and BMPR2, CREB1, KRAS, MAP3K7, PRKAR2B, SMAD5, XIAP (7 of 8 BMP pathway members) are significantly (**P<0.008) down-regulated in cluster 1 compared to cluster 2 (Mann-Whitney U, two tailed, 95% confidence interval). Expression data are displayed as mean±standard deviation. Glomerular FRMD3 expression cluster 1: 8.29±0.54, cluster 2: 9.67±0.41.

Fig. 2 D: FRMD3/BMP repression is associated with increase of albuminuria.
Clinical measures of delta ACR/year comparing the 2 main cluster branches from Fig. 2B. Data are displayed as mean±standard deviation. Delta ACR/year cluster 1: 212.4±227.9 is significantly (*P=0.017) increased.
compared to Delta ACR/year in cluster 2: 3.7±8.7. (Mann-Whitney U, two tailed, 95% confidence interval).

Fig. 2 E: FRMD3/BMP repression is associated with increase of fractional mesangial area. Histologic measures of fractional mesangial area (%) comparing the 2 main cluster branches from Fig. 2B. Data are displayed as mean±standard deviation. Mesangial expansion was significantly (*P=0.04) increased in cluster 1 (30±14%) compared to cluster 2 (17±7%) (Mann-Whitney U, two tailed, 95% confidence interval).
Fig. 3: Functional association of FRMD3-correlated genes. Top 10 pathways (Ingenuity Pathways Analysis, Ingenuity® Systems) of 581 FRMD3-correlated genes sorted by the ratio of members of the pathway among FRMD3-correlated genes vs. total number of members of that pathway. P ≤ 0.001 for all pathways.
Fig. 4: Increased binding of glomerular nuclear extracts to DN-associated genomic region. Electrophoretic mobility shift assay (EMSA) from oligonucleotides corresponding to the wild type DNA sequence (WT) and SNP-altered sequence (SNP) of glomerular extracts from C57Black6 mice. The nonspecific competitor poly dIdC was used. Arrow indicates position of protein-bound oligos. With increasing amounts of protein used a distinct binding signal can be detected in the SNP-sequence but to a lesser amount in the WT-sequence as displayed in the Intensity Blot B). Intensity of the DNA-protein complex in lane 2 was set to 1.0. A paired t-test showed that the mean intensity was significantly higher in the SNP sequences compared to the WT-sequence (P=0.04).

228x291mm (300 x 300 DPI)
Suppl. Table 1: Transcripts with mRNA expression correlation to *FRMD3* in DN.
Co-regulated transcripts listed by descending person correlation. Top 10 transcripts are concordantly regulated in DN. For 5 of the 10 transcripts or close variants an association with diabetes, cardiovascular pathophysiology or inflammatory diseases has been published, summarized in columns 4 and 5.

Suppl. Table 2: Clinical characteristics of 22 Pima Indians with T2D diabetic nephropathy
Age at the time of biopsy, gender, iothalamate clearance (iGFR) at the time of biopsy, ACR at the time of biopsy, delta GFR (mL/min/year), delta ACR/year, fractional mesangial area (%), observation time (defined as time of enrollment until latest available clinical data point) and *FRMD3* steady state mRNA levels are shown as mean group values±standard deviation (SDEV) where suitable for both main clusters. Subjects are sorted by affiliation to one of the main clusters displayed in Fig. 1. *P<0.05 and **P<0.01 between cluster 1 and cluster 2.
<table>
<thead>
<tr>
<th>Gene Name</th>
<th>Gene ID</th>
<th>R with FRMD3</th>
<th>Relevance for diabetes, cardiovascular/inflammational disease</th>
<th>PMID</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADD3</td>
<td>120</td>
<td>0.93</td>
<td>interaction of ADD1 and ADD3 gene variants in humans is statistically associated with variation in blood pressure</td>
<td>15716695</td>
</tr>
<tr>
<td>DNAJB14</td>
<td>79982</td>
<td>0.93</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DACH1</td>
<td>1602</td>
<td>0.9</td>
<td>differentially regulated in diabetic mouse model, candidate gene identified by Diabetes Genetics Initiative and Wellcome Trust Case Control Consortium genome-wide scans for T2D</td>
<td>19902174</td>
</tr>
<tr>
<td>ZYG11B</td>
<td>79699</td>
<td>0.9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FKBP14</td>
<td>55033</td>
<td>0.89</td>
<td>FKB12.6-knockout mice display hyperinsulinemia and resistance to high-fat diet-induced hyperglycemia</td>
<td>19805579</td>
</tr>
<tr>
<td>Gene</td>
<td>NCBI ID</td>
<td>Score</td>
<td>Description</td>
<td>Reference Numbers</td>
</tr>
<tr>
<td>------------</td>
<td>---------</td>
<td>-------</td>
<td>-----------------------------------------------------------------------------</td>
<td>-------------------</td>
</tr>
<tr>
<td>FAM126B</td>
<td>285172</td>
<td>0.88</td>
<td></td>
<td></td>
</tr>
<tr>
<td>KIF5B</td>
<td>3799</td>
<td>0.88</td>
<td>essential for cardiac potassium channels, candidate gene for the response of cardiac stroke volume</td>
<td>19675065, 18984674</td>
</tr>
<tr>
<td>LOC153222</td>
<td>153222</td>
<td>0.88</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LYRM7</td>
<td>90624</td>
<td>0.88</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RSBN1</td>
<td>54665</td>
<td>0.88</td>
<td>genetic variant in RSBN1 locus associated with juvenile idiopathic arthritis</td>
<td>20722033</td>
</tr>
<tr>
<td>Cluster</td>
<td>Age</td>
<td>Gender</td>
<td>GFR (mL/min)</td>
<td>ACR</td>
</tr>
<tr>
<td>---------</td>
<td>-----</td>
<td>--------</td>
<td>--------------</td>
<td>-----</td>
</tr>
<tr>
<td>1</td>
<td>44±15</td>
<td>4xF 2xM</td>
<td>148±37</td>
<td>584±932</td>
</tr>
<tr>
<td>2</td>
<td>46±10</td>
<td>14xF 2xM</td>
<td>132±42</td>
<td>48±53</td>
</tr>
</tbody>
</table>