Uncovering Genetic Factors Contributing to Type 2 Diabetes and Diabetic Nephropathy

Y. Bai,[1] V. Nair,[2] S. Martini,[2] J. Cavalcoli,[1] G. Omenn,[1] M. Kretzler[1,2]

[1] National Center for Integrative Biomedical Informatics, and Center for Computational Medicine and Biology; [2] Department of Internal Medicine
University of Michigan, Ann Arbor, MI 48109

Introduction

Recent genome-wide association studies [1-4] have made major discoveries in identifying Type 2 diabetes (T2D) associated regions and loci, but the specific sequence variances responsible for the associations remain elusive.

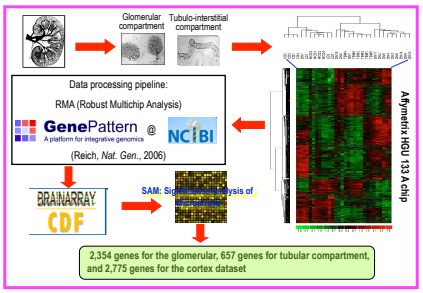
To define putative causative gene sets from GWAS we employed a promoter modeling approach based on the hypothesis that promoter regions integrate upstream signaling cascades towards coordinated transcription of functionally interdependent mRNAs. Defining T2D dependent promoter models in GWAS candidate promoters might thereby facilitate identification of putative causative transcript alterations.

Here we studied the proximal promoter regions of 13 genes selected from T2D associated regions in the 3-way FUSION-DGI-WTCCC meta-analysis [1, 2, 4]. In particular, we used computational methods to identify shared putative regulatory promoter modules in the proximal promoter regions that we investigated. Specific potential regulatory promoter modules containing three transcription factor (TF) binding motifs in a defined order and spacing were identified in a subset of genes chosen from GWAS associated regions. These promoter modules helped elucidate other module sharing genes in the GWAS, which are possibly regulated in a similar fashion.

Our study provides TF binding module data that can putatively activate a subset of T2D GWAS genes.

Methods

1. Selecting significantly regulated transcripts from Diabetic Endorgan Damage (Diabetic Nephropathy) and module genes from T2D GWAS
2. Feeding selected genes into TFBS/promoter module analysis pipeline [6]
3. Training test T2D GWAS genes for identified promoter modules

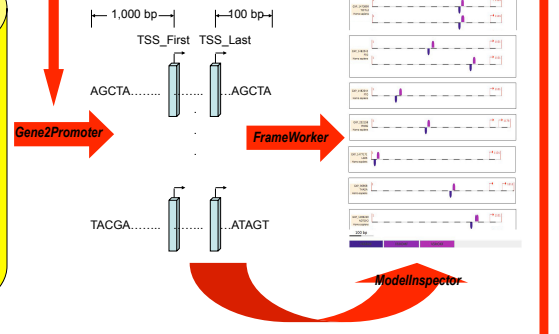


1

The filtered PIMA gene expression dataset resulted in ~11000 genes to be expressed above background. 21 out of the 29 T2D associated genes were mapped to this dataset. SAM analysis on these 21 genes in the three datasets (glomeruli, tubulointerstitium and cortex) resulted in 12 significantly regulated genes. We additionally incorporated "CDKN2B" into the input gene list.

2

PPARG
CDC123
CAMK1D
TSPAN8
LGR5
THADA
NOTCH2
IGF2BP2
CDKN2A
CDKN2B
TCF7L2
HHEX
FTO



3

We chose a region within 200kb of each SNP with p<.001 in the 3-way FUSION_DGI_WTCCC meta-analysis [1, 2, 4]. Two additional genes from large scale studies were added: Wolfram Syndrome (WFS1) and HNF1B.

13 module genes in the training set

34 additional genes from published regions

3 genes overlapped

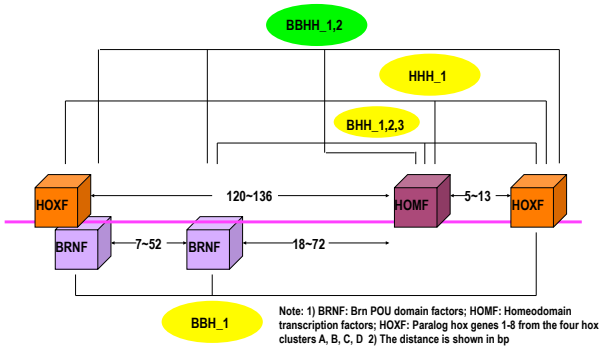
17 genes sharing 4 TF modules that we have identified

171 other genes selected from 3-way T2D meta-analysis (P <.0001 and within 200kb of SNPs)

48 TEST genes

Results

➢ **Initial modules identified**



Note: 1) BRNF: Brn POU domain factors; HOMF: Homeodomain transcription factors; HOXF: Paralog hox genes 1-8 from the four hox clusters A, B, C, D 2) The distance is shown in bp

➢ **TF binding modules enrichment in 48 selected genes**

Model Name[a]	Genome-wide: Total # of promoter sequences in human genome	Genome-wide: # of module presence in human promoters[c]	Genome-wide: Ratio	48 GWAS Genes: Total # of promoter sequences in 48 GWAS genes	48 GWAS Genes: # of module presence in the promoters of 48 GWAS genes	48 GWAS Genes: Ratio	Fold change (48 GWAS genes/Genome-wide[d])	Z Score[e]	Fisher's exact test[f]
BBH_1	95606[b]	13135	0.14	260	67	0.26	1.88	5.62	2.91×10^{-7}
BBO_1	95606	5379	0.06	260	28	0.11	1.91	3.59	0.001
BHH_1_2	95606	27898	0.29	260	152	0.58	2.00	10.36	1.42×10^{-22}
BHH_3	95606	12221	0.13	260	57	0.22	1.72	4.40	5.32×10^{-5}
HHH_1	95606	8256	0.09	260	46	0.18	2.05	5.19	3.78×10^{-6}

[a] Model Name Abbreviation: BRNF_BRNF_HOXF_1 (BBH_1); BRNF_BRNF_OCT1_1 (BBO_1); BRNF_HOMF_HOXF_1 and BRNF_HOMF_HOXF_2 (BHH_1_2); BRNF_HOMF_HOXF_3 (BHH_3); HOXF_HOMF_HOXF_1 (HHH_1)
[b] Bolded models are significant
[c] Results are based on Genomatix Module Library 5.0 and Matrix Library 7.1
[d] Fold change >1.5 is considered significant
[e] The Z-score is calculated as the distance from the population mean in units of thepopulation standard deviation. A Z-score above 2 is considered statistically significant
[f] Fisher exact test results are based on promoter sequences

Future work

- Develop methods for incorporating the interactions for TF factors linked to putative module genes into cellular functional context via NCIBI-MiMI database
- Run NCIBI-ConceptGen software on 61 genes
- Apply NCIBI tools (MiMI, Gene2MeSH, and SAGA) to look for detailed protein interactions, possible MeSH term enrichment, and pathway involvement for a subset of genes which share four TF binding modules to define functional context of co-regulation

References

1. Scott, L.J., et al., *A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants.* Science, 2007. **316**(5829): p. 1341-5.
2. Saxena, R., et al., *Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels.* Science, 2007. **316**(5829): p. 1331-6.
3. Zeggini, E., et al., *Meta-analysis of genome-wide association data and large-scale replication identifies additional susceptibility loci for type 2 diabetes.* Nat Genet, 2008. **40**(5): p. 638-45.
4. *Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls.* Nature, 2007. **447**(7145): p. 661-78.
5. Reich, M., et al., *GenePattern 2.0.* Nat Genet, 2006. **38**(5): p. 500-1.
6. Scherf, M., A. Klingenhoff, and T. Werner, *Highly specific localization of promoter regions in large genomic sequences by PromoterInspector: a novel context analysis approach.* J Mol Biol, 2000. **297**(3): p. 599-606.

Acknowledgement

This work was supported by National Institutes of Health: Grant #U54 DA021519