Carriers of a VEGFA enhancer polymorphism selectively binding chop/ddit3 are predisposed to increased circulating levels of thyroid stimulating hormone

Ahluwalia, Tarunveer Singh; Troelsen, Jesper; Balslev-Harder, Marie; Bork-Jensen, Jette; Thuesen, Betina Heinsbæk; Cerqueira, Charlotte; Linneberg, Allan; Grarup, Niels; Pedersen, Oluf; Hansen, Torben; Dalgaard, Louise Torp

Published in:
Journal of Medical Genetics

DOI:
10.1136/jmedgenet-2016-104084

Publication date:
2017

Document Version
Peer reviewed version

Citation for published version (APA):
Ahluwalia, T. S., Troelsen, J., Balslev-Harder, M., Bork-Jensen, J., Thuesen, B. H., Cerqueira, C., ... Dalgaard, L. T. (2017). Carriers of a VEGFA enhancer polymorphism selectively binding chop/ddit3 are predisposed to increased circulating levels of thyroid stimulating hormone. Journal of Medical Genetics, 54(3), 166-175. DOI: 10.1136/jmedgenet-2016-104084

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain.
• You may freely distribute the URL identifying the publication in the public portal.

Take down policy
If you believe that this document breaches copyright please contact rucforsk@ruc.dk providing details, and we will remove access to the work immediately and investigate your claim.

Download date: 15. Jan. 2019
CARRIERS OF A VEGFA ENHANCER POLYMORPHISM SELECTIVELY BINDING CHOP/DDIT3 ARE PREDISPOSED TO INCREASED CIRCULATING LEVELS OF THYROID STIMULATING HORMONE

<table>
<thead>
<tr>
<th>Journal:</th>
<th>Journal of Medical Genetics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manuscript ID</td>
<td>jmedgenet-2016-104084.R1</td>
</tr>
<tr>
<td>Article Type:</td>
<td>Original Article</td>
</tr>
<tr>
<td>Date Submitted by the Author:</td>
<td>n/a</td>
</tr>
</tbody>
</table>
| Complete List of Authors: | Ahluwalia, Tarunveer; University of Copenhagen, The Novo Nordisk Foundation Center for Basic Metabolic Research, Section of Metabolic Genetics, Faculty of Health and Medical Sciences; Steno Diabetes Center AS  
Troelsen, Jesper; Roskilde Universitet, Department of Science and Environment  
Balslev-Harder, Marie; University of Copenhagen, 1The Novo Nordisk Foundation Center for Basic Metabolic Research, Section of Metabolic Genetics, Faculty of Health and Medical Sciences  
Bork-Jensen, Jette; University of Copenhagen, The Novo Nordisk Foundation Center for Basic Metabolic Research, Section of Metabolic Genetics, Faculty of Health and Medical Sciences  
Thuesen, Betina; Research Centre for Prevention and Health  
Cerqueira, Charlotte; Research Centre for Prevention and Health  
Linneberg, Allan; Research Centre for Prevention and Health  
Grarup, Niels; Kobenhavns Universitet Sundhedsvidsenskabelige Fakultet, The Novo Nordisk Foundation Center for Basic Metabolic Research  
Pedersen, Oluf; Kobenhavns Universitet Sundhedsvidsenskabelige Fakultet, The Novo Nordisk Foundation Center for Basic Metabolic Research  
Hansen, Torben; Kobenhavns Universitet Sundhedsvidsenskabelige Fakultet, The Novo Nordisk Foundation Center for Basic Metabolic Research  
Dalgaard, Louise; Roskilde Universitet, Department of Science and Environment |
| Keywords:         | thyroid, Molecular genetics, Metabolic disorders, insulin resistance, transcription factor |

https://mc.manuscriptcentral.com/jmedgenet
CARRIERS OF A VEGFA ENHANCER POLYMORPHISM SELECTIVELY BINDING CHOP/DDIT3 ARE PREDISPOSED TO INCREASED CIRCULATING LEVELS OF THYROID STIMULATING HORMONE

Running title: Circulating TSH association with a VEGFA functional polymorphism

Tarunveer Singh Ahluwalia, Jesper Thorvald Troelsen, Marie Balslev-Harder, Jette Bork-Jensen, Betina Heinsbæk Thuesen, Charlotte Cerqueira, Allan Linneberg, Niels Grarup, Oluf Pedersen, Torben Hansen, Louise Torp Dalgaard

1 The Novo Nordisk Foundation Center for Basic Metabolic Research, Section of Metabolic Genetics, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark
2 Steno Diabetes Center, Gentofte, Denmark
3 Department of Science and Environment, Roskilde University, Roskilde, Denmark
4 Research Centre for Prevention and Health, The Capital Region, Glostrup, Denmark
5 Department of Clinical Experimental Research, Rigshospitalet, Copenhagen, Denmark
6 Department of Clinical Medicine, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark
7 Faculty of Health Sciences, University of Southern Denmark, Odense, Denmark

Address correspondence to:
Associate professor Louise Torp Dalgaard, MSc, PhD, Department of Science and Environment, Roskilde University, Universitetsvej 1, DK-4000 Roskilde, Denmark. E-mail: ltd@ruc.dk, phone: +45 4674 2713

Figures: 4, Supplementary figures: 4

KEYWORDS
thyroid, genetics, metabolic disorders, insulin resistance, transcription factor
ABSTRACT

Background: Levels of serum thyroid stimulating hormone (TSH) indicate thyroid function, because thyroid hormone negatively controls TSH release. Genetic variants in the vascular endothelial growth factor A (VEGFA) gene are associated with TSH levels. The aim was to characterize the association of VEGFA variants with TSH in a Danish cohort and to identify and characterize functional variants.

Methods: We performed an association study of the VEGFA locus for circulating TSH levels in 8445 Danish individuals. Lead variants were tested for allele-specific effects in vitro using luciferase reporter and gel-shift assays.

Results: Four SNPs in VEGFA were associated with circulating TSH (rs9472138, rs881858, rs943080 and rs4711751). For rs881858, the presence of each G allele was associated with a corresponding decrease in TSH levels of 2.3% (P=8.4×10⁻⁹) and an increase in circulating free T4 levels (P=0.0014). Rs881858 is located in a binding site for CHOP (C/EBP homology protein) and C/EBPβ (ccaat enhancer binding protein β). Reporter-gene analysis showed increased basal enhancer activity of the rs881858 A-allele versus the G-allele (34.5±9.9% (average±SEM), P=0.0012), while co-expression of CHOP effectively suppressed the rs881858 A-allele activity. The A-allele showed stronger binding to CHOP in gel-shift assays.

Conclusions: VEGF is an important angiogenic signal required for tissue expansion. We show that VEGFA variation giving allele-specific response to transcription factors with overlapping binding sites associate closely with circulating TSH levels. Because CHOP is induced by several types of intracellular stress, this indicates that cellular stress could be involved in the normal or pathophysiological response of the thyroid to TSH.
ABBREVIATIONS

TSH thyroid stimulating hormone, VEGFA vascular endothelial growth factor A, CHOP c/EBP homology protein, c/EBPβ ccaat enhancer binding protein β, CEBPB C/EBPβ gene symbol, SNP single nucleotide polymorphism, Chr chromosome, EAF effect allele frequency, BS binding site, ds double-stranded, DTT dithiothreitol, GWAS genome wide association study, BMI body mass index, eQTL expressed quantitative trait locus
INTRODUCTION

The thyroid gland is an essential regulator whole body energy expenditure and metabolic rate. Circulating levels and activities of thyroid hormones, their activating enzymes (deiodinases) and the regulating hormones TSH (thyrotropin/thyroid stimulating hormone) and TRH (thyrotropin releasing hormone) are precisely balanced to ensure the euthyroid state. Circulating levels of TSH comprise a clinically valuable indicator of thyroid function, and in the absence of pituitary or hypothalamic failure, an increased level of TSH is a very sensitive marker of decreased thyroid function. Clinical reference levels for TSH define elevated TSH levels above 4mU/L as being associated with clinically decreased thyroid function\(^1\). However, there is a marked inter-individual, while low intra-individual variability in circulating TSH levels, as well as in the hypothalamic-pituitary-thyroid axis\(^2\), which appears to be highly heritable as evidenced by heritability estimates of 65% derived from twin studies\(^3-5\).

Subclinical hypothyroidism (TSH above 4mU/L with T4 levels within the reference range) is associated with an impaired metabolic phenotype, cardiovascular risk factors, elevated blood total cholesterol and blood pressure increase, decreased glomerular filtration rate and bone fractures\(^6\). However, association between TSH within the reference range and obesity is not well established. A meta-analysis encompassing 29 studies, found 18 of these to report a positive association\(^7\). There have been reports of positive correlations between TSH levels and BMI in obese or over-weight individuals, suggesting decreased thyroid function in these subjects\(^8-10\), but also cross sectional population-based reports have been made for the association between increased TSH levels and BMI\(^11-13\).

Recent genome wide association studies (GWAS) for circulating serum TSH levels have focused on common variants (minor allele frequency (MAF) > 5%) and have identified at least 26 genomic loci to date\(^14-17\) of which one was the Vascular Endothelial Growth Factor A gene (VEGFA) (Fig. 1A).
Both the VEGF protein and its receptor, KDR (Kinase Insert Domain Receptor), are highly expressed in the thyroid gland\textsuperscript{18}. \textit{In vivo}, TSH administration increases VEGF release from the thyroid gland\textsuperscript{19} and treatment of isolated, cultured thyrocytes with TSH also stimulates VEGF release\textsuperscript{20}. Variants regulating the activity of the \textit{VEGFA} locus are likely contributors to the observed TSH association\textsuperscript{14}, because TSH levels indicate thyroid function.

In order to investigate the molecular and genetic mechanisms in the \textit{VEGFA} locus controlling circulating levels of TSH, we performed an association study using densely spaced SNPs of the \textit{VEGFA} genomic region in three population-based cohorts from Denmark comprising 8,445 individuals and meta-analysed them. The lead variants were further investigated to identify allele specific effects using \textit{in vitro} cell based assays to elucidate the molecular mechanism supporting the observed clinical findings.

**METHODS**

**Genetic association analysis**

\textit{Study participants}

The study was conducted in accordance with the Helsinki declaration and approved by the Danish Data Protection Board and by the Ethical Committee of Copenhagen County. Informed written consent was obtained from all subjects before participation. The genetic association analyses were performed in three Danish study cohorts (Inter99, Health2006 and Health2008) that have been described previously elsewhere\textsuperscript{21}: 1) The study (ClinicalTrials.gov ID-no: NCT00289237) is a population-based study for ischemic heart disease\textsuperscript{22}, 2) The Health 2006 Study (Ethical committee approval number: KA20060011) is a population based study comprising individuals aged between 18-69 years from the South Western part of greater Copenhagen area\textsuperscript{23}. This study was designed to address chronic health issues, 3) The Health 2008 Study (Ethical committee approval number: KA20060011) is a cross sectional population based study\textsuperscript{24}. 
Biochemical and anthropometric measurements

The biochemical and anthropometric information on and phenotypical characterization of study participants is described in Table S1 and Table S2 and has been presented previously\textsuperscript{22-24}.

Genotyping, variant calling and quality checks

DNA extraction, genotyping and genotype call processing has been described previously\textsuperscript{21 25}. A total of 8,445 individuals (n\textsubscript{Intergen99}:5,420, n\textsubscript{Health2006}:2,442, n\textsubscript{Health2008}:583) with complete phenotype and genotype data participated for the serum TSH association analyses. The genotyping platform was Illumina Human Exome 12v1.0 containing 263.894 SNPs (including 16.024 custom SNPs identified from a recent exome sequencing study in Danes\textsuperscript{21 25}) post quality control.

SNP Selection

SNPs from the VEGFA gene region (6:43737946-6:43754224 GRCh37) and within the adjoining flanking region (±75 Kb) were selected covering a total of 166.2 kb (6:43662946-6:43829224 GRCh37) region. After removing SNPs with a MAF<0.005 a total of 15 non-coding SNPs were available from the VEGFA gene region for this study. Based on this SNP selection, a Bonferroni corrected \( p \) value corresponding 0.0033 was set as the significance threshold for SNP-TSH association testing.

SNP-TSH association testing and meta-analyses

The association testing between the SNPs and the serum TSH levels was performed individually for each cohort using the additive linear regression model adjusting for gender, age and first five principal components as covariates. The fasting circulating measures of TSH were transformed to natural log scale before the association testing to control for non-normalised data. Prior to the association testing, individuals with known thyroid pathologies and those with out of range TSH values (<0.4 mU/L and >4.0 mU/L) were removed. The method for meta-analyses was as previously described\textsuperscript{21}.

SNP metabolic traits association
Associations between a SNP and metabolic traits were tested among normal glucose tolerant (NGT) individuals from the Inter99 cohort using the general linear model assuming an additive genetic effect for the SNP. Association was tested with baseline measures and changes during follow-up ($\Delta$ values: *Follow up – baseline measurements*) and was adjusted for gender and age. A $p$-value of $p<0.05$ was considered significant for the follow-up and single SNP-metabolic trait testing. All association analyses were performed using PLINK v1.07 (http://pngu.mgh.harvard.edu/purcell/plink/), and R version 3.1.1 (http://www.r-project.org/).

Linkage Disequilibria (LD) estimations

LD estimation and proxy search was performed using 1000 genomes project data and an LD heat-map depicting pairwise $r^2$ values is available in Fig. S1. Non-genotyped SNPs in LD with TSH-associated SNPs were retrieved from the ENSEMBL genome browser and SNPs with LD≥0.8 were inspected for evidence of differential allele effects using Genome Browser (https://genome.ucsc.edu/) (Table S3).

In vitro molecular biology studies

Reporter gene analysis

The *VEGFA* proximal promoter (hg19, chr6:43737097-43738057), corresponding to 849bp of promoter and 252bp of the first intron, was cloned into pGL4.10. Genomic regions containing rs881858 and rs9472138 were amplified by PCR from homozygous carriers and cloned into *VEGFA* pro/GL4.10 down-stream of the *luc* gene to generate rs88 A/GL4.10, rs88 G/GL4.10, rs94 C/GL4.10 and rs94 T/GL4.10. All constructs were confirmed by sequencing. Plasmid DNA was prepared using Qiagen Maxi Prep kit (Qiagen, Copenhagen Ø, Denmark) and ethanol precipitated. Transfections were made in human embryonic kidney cells (HEK)-293 cells (American Tissue Type Culture Collection, Rockville, MD, U.S.A.) using polyethylene imine (PEI25). Cells were harvested after 24hrs for luciferase assays (Dual Light, Thermofisher Scientific, Copenhagen Ø, Denmark).
Expression vectors for C/EBPβ and CHOP10 were a gift from Peter Johnson (Addgene plasmid #12557) and David Ron (Addgene plasmid #21899), respectively.

**Electrophoretic mobility shift assay (EMSA)**

Nuclear extracts from HEK293 cells were prepared as described previously. Some extracts were prepared following incubation of cells with 1mM dithiothreitol (DTT) for 16hrs to induce CHOP10 via the unfolded protein response. Complementary oligos representing the SNPs rs881858 were annealed and labelled with α-32P-dATP (3000 Ci/mmol) by Klenow fill-in and purified using NICK columns (GE Healthcare, Brøndby, Denmark). Binding reactions were made as described previously, separated with non-denaturing polyacrylamide gel-electrophoresis and visualized on phosphor-imager screen. Screens were scanned using a Molecular Dynamics Storm Scanner and the protein/DNA complexes analyzed using Image-Quant Software version 3.5. Oligonucleotides used for cloning and EMSA are listed in Table S3.

**Statistics for molecular biology experiments**

Results are expressed as mean value ± SEM. Statistical analysis was performed in GraphPad Prism software. Effects of SNP constructs were tested using ANOVA with post-hoc t-test and Bonferroni correction. Differences between treatments were considered significant at a P-value<0.05 (two-tailed).

**RESULTS**

**VEGFA and TSH association analyses**

We investigated three population-based Danish cohort studies for association analyses of the VEGFA locus with circulating TSH: The Inter99, the Health2006 and Health2008 cohorts (Table S1). We searched for genotyped SNPs located up to 75kb up- and downstream of the VEGFA transcription start site (TSS) (VEGFA: Chr6:43737946-43754224). This region was defined based on the localization of chromatin marks within this region (Chromatin Interaction Analysis by paired-end
sequencing, ChIA-PET indicating interaction with the VEGFA gene and localization of nearby transcripts MRPS18A (upstream of VEGFA) and LOC100132354 and C6orf223 downstream of VEGFA) and we identified 15 SNPs in this region (Table 1, Table S4).
Table 1: VEGFA SNPs significantly associated with circulating levels of thyroid stimulating hormone (TSH)

<table>
<thead>
<tr>
<th>SNP name</th>
<th>Position (build 37/hg19)</th>
<th>Location wrt VEGFA</th>
<th>Alleles (effect/other)</th>
<th>EAF</th>
<th>Inter99 n=5,420</th>
<th>Health2006 n=2,442</th>
<th>Health2008 n=583</th>
<th>Combined</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Effect</td>
<td>P</td>
<td>Effect</td>
<td>P</td>
</tr>
<tr>
<td>rs9472138</td>
<td>43,811,762</td>
<td>Downstream</td>
<td>T/C</td>
<td>0.27</td>
<td>-0.046</td>
<td>(4.4 \times 10^{-6})</td>
<td>-0.051</td>
<td>(1.0 \times 10^{-5})</td>
</tr>
<tr>
<td>rs881858</td>
<td>43,806,609</td>
<td>Downstream</td>
<td>G/A</td>
<td>0.28</td>
<td>-0.046</td>
<td>(2.9 \times 10^{-6})</td>
<td>-0.047</td>
<td>(1.9 \times 10^{-5})</td>
</tr>
<tr>
<td>rs943080</td>
<td>43,826,672</td>
<td>Downstream</td>
<td>T/C</td>
<td>0.51</td>
<td>-0.024</td>
<td>0.0053</td>
<td>-0.012</td>
<td>0.35</td>
</tr>
<tr>
<td>rs4711751</td>
<td>43,828,582</td>
<td>Downstream</td>
<td>T/C</td>
<td>0.51</td>
<td>-0.023</td>
<td>0.0089</td>
<td>-0.012</td>
<td>0.32</td>
</tr>
</tbody>
</table>

*SNPs in LD \((r^2>0.4)\). **SNPs in strong LD \((r^2: 0.73, D':0.94)\). ***SNPs in strong LD \((r^2: 1.0, D': 1.0)\). EAF: Effect allele frequency. VEGFA: Vascular endothelial growth factor A. \(F\): heterozygosity at meta-analyses level. \(P_{Het}\): P value for heterozygosity. List of SNPs not reaching study wide association with circulating TSH-levels is given in Suppl. Table S4.
Four VEGFA region SNPs (rs9472138, rs881858, rs943080, and rs4711751) associated with fasting serum TSH levels at a study-wide significance level ($P_{combined} < 0.0033$), (Table 1, Fig. 1A and Fig. S2) in up to 8445 individuals following combined-meta analysis of the three cohorts. All four SNPs were in LD with each other ($r^2 > 0.4$), and one SNP was a known signal (VEGFA rs9472138)\textsuperscript{14} for circulating TSH. All the significantly associating SNPs were common (MAF>0.05).

**Genomic marks qualifying VEGFA SNPs for further investigation**

Common SNPs in high LD ($r^2 >0.8$) (proxy SNPs) with the four significant SNPs in VEGFA: rs9472138, rs881858, rs943080 and rs4711751 was obtained from the ENSEMBL browser yielding 12 common SNPs linked with rs9472138 and rs881858 and 3 SNPs with rs943080 and rs4711751 (Table S5, Fig. S1), previously genotyped in the 1000genomes project. These SNPs were considered functional candidates to explain the observed genetic association, because they are common and in high LD with the lead SNPs. Using ENCODE data-tracks on Genome Browser (GRCh37/hg19\textsuperscript{28} these SNPs were evaluated based on presence of open chromatin structure (DNase seq), conservation, marks of H3K27Ac (Histone 3, Lysine 27 acetylation) and H3K4Me1 (Histone 3, Lysine 4 mono-methylation) and indication of protein binding to the SNP region by ChIP-seq. Furthermore, it was also assessed if the SNP altered the binding site for factors shown to bind the region by ChIP (Table S5, Fig.1B). From this, it was evident that rs881858 was highly conserved with G being the ancestral allele. Furthermore, rs881858 was located in a region showing marks characteristic of active regulatory elements (H3K27Ac and H3K4Me1 and displaying evidence of close three dimensional proximity to the VEGFA promoter (Chromosome Interaction Analysis-Paired End Tags (ChIA-PET). The region was also DNase hypersensitive suggesting an open chromatin structure (Fig. 1B).

Moreover, rs881858 was directly located in a site bound by the transcription factor c/EBPβ (encoded by CEBPB) (by ChIP-seq) in multiple cell lines, and predicted to be bound by CHOP. Thus, for rs881858 there is strong evidence of an allele-specific regulatory role. Performing the same analysis
for the other SNPs did not reveal equal evidence of regulatory activity, conservation or protein
binding to their vicinity (Table S5, Fig. S3). Rs881858 is located at the 3’ end of the c/EBPβ binding
site BS in a position, which does not confer specificity to the c/EBPβ binding according to the
position weight matrix (PWM) for c/EBPβ (Fig. 1C), whereas the BS for CHOP is predicted to
prefer the A-allele of the rs881858 compared with the G-allele. Thus, based on ENSEMBL and
ENCODE data and differential predicted binding affinities to c/EBPβ and CHOP, the A and G alleles
of rs881858 could confer differential responses.

**Association of G-allele of VEGFA rs881858 with decreased circulating TSH, increased thyroid
hormone levels and metabolic traits**

The effect size of rs881858 on circulating TSH levels was -0.092 (P=2.2×10^{-11}) for the G-allele
corresponding to an additive 2.3% decrease in TSH level per allele. Furthermore, free T4 levels were
correspondingly increased in GG subjects compared with AA or AG subjects (GG: 15.4 (14.2-16.6)
vs AA: 15.0 (13.8-16.3) pmol/L, P=0.0014), indicating a slightly increased thyroid function in GG
individuals and an altered set point for the TSH/T4 axis (Table 2). There was no available
information on circulating levels of total T3 or thyroid hormone binding globulin for these subjects.
Table 2: Associations between VEGFA rs881858 G-allele and metabolic traits in 4,373 normal glucose tolerant (NGT) Danish subjects (Inter99) at baseline

<table>
<thead>
<tr>
<th>VEGFA rs881858</th>
<th>GG</th>
<th>GA</th>
<th>AA</th>
<th>*Effect</th>
<th>SE</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>n</td>
<td>347</td>
<td>1,754</td>
<td>2,272</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Age (years)</td>
<td>45 (40-50)</td>
<td>45 (40-50)</td>
<td>45 (40-50)</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>TSH (mIU/L)</td>
<td>1.11 (0.77-1.56)</td>
<td>1.23 (0.88-1.75)</td>
<td>1.29 (0.90-1.85)</td>
<td>-0.092</td>
<td>0.013</td>
<td>2.2 \times 10^{-11}</td>
</tr>
<tr>
<td>Free T4 (pmol/L)</td>
<td>15.4 (14.2-16.6)</td>
<td>15.0 (13.9-16.3)</td>
<td>15.0 (13.8-16.3)</td>
<td>0.009</td>
<td>0.002</td>
<td>1.4 \times 10^{-3}</td>
</tr>
<tr>
<td>BMI (kg/m²)</td>
<td>25.1 (22.8-28.2)</td>
<td>24.9 (22.8-27.6)</td>
<td>24.9 (22.7-27.5)</td>
<td>0.14</td>
<td>0.095</td>
<td>0.11</td>
</tr>
<tr>
<td>Leptin (ng/ml)</td>
<td>5.5 (2.5-10.9)</td>
<td>5.3 (2.4-11.0)</td>
<td>5.3 (2.4-11.3)</td>
<td>0.01</td>
<td>0.022</td>
<td>0.62</td>
</tr>
<tr>
<td>HbA1c (%)</td>
<td>5.8 (5.5-6.1)</td>
<td>5.8 (5.5-6.0)</td>
<td>5.8 (5.5-6.0)</td>
<td>0.016</td>
<td>0.008</td>
<td>0.056</td>
</tr>
<tr>
<td>Fasting plasma glucose (mmol/L)</td>
<td>5.4 (5.1-5.6)</td>
<td>5.3 (5.0-5.6)</td>
<td>5.3 (5.0-5.6)</td>
<td>0.021</td>
<td>0.008</td>
<td>0.016</td>
</tr>
<tr>
<td>2-hour glucose during OGTT (mmol/L)</td>
<td>5.6 (4.8-6.4)</td>
<td>5.6 (4.7-6.4)</td>
<td>5.5 (4.7-6.3)</td>
<td>0.072</td>
<td>0.026</td>
<td>6.1 \times 10^{-3}</td>
</tr>
<tr>
<td>Fasting serum insulin (pmol/L)</td>
<td>32 (23-47.5)</td>
<td>31 (22-45)</td>
<td>31 (21-44)</td>
<td>0.023</td>
<td>0.013</td>
<td>0.073</td>
</tr>
<tr>
<td>2-hour insulin during OGTT (pmol/L)</td>
<td>133 (86-221)</td>
<td>131 (83-202)</td>
<td>134 (83-204)</td>
<td>0.013</td>
<td>0.017</td>
<td>0.43</td>
</tr>
<tr>
<td>ISI Matsuda</td>
<td>3.03 (2.2-4.2)</td>
<td>3.15 (2.24-4.44)</td>
<td>3.19 (2.25-4.47)</td>
<td>-0.02</td>
<td>0.012</td>
<td>0.12</td>
</tr>
<tr>
<td>HOMA-IR</td>
<td>1.25 (0.92-1.92)</td>
<td>1.20 (0.83-1.81)</td>
<td>1.20 (0.82-1.76)</td>
<td>0.027</td>
<td>0.013</td>
<td>0.041</td>
</tr>
<tr>
<td>Insulinogenic index</td>
<td>85.2 (53.3-137.2)</td>
<td>76.0 (47.6-125.1)</td>
<td>77.0 (49.7-129.2)</td>
<td>0.013</td>
<td>0.017</td>
<td>0.45</td>
</tr>
<tr>
<td>Disposition index</td>
<td>227.0 (169.3-339.8)</td>
<td>226.0 (159.6-341.0)</td>
<td>231.6 (162.7-352.7)</td>
<td>-0.006</td>
<td>0.015</td>
<td>0.68</td>
</tr>
</tbody>
</table>

Values correspond to median (interquartile range) in non-transformed traits. * G allele as the effect allele assuming an additive genetic model for log transformed traits. ISI Matsuda, HOMA-IR, Insulinogenic index and Disposition index were calculated as described in Supplementary Table S3. P values in bold indicate significant results (P<0.05).

Since increased circulating TSH levels are associated with an impaired metabolic phenotype, it was tested if rs881858 associated with measures of glucose tolerance. The GG genotype of rs881858 was associated with slightly increased fasting and 2-hour post OGTT plasma glucose levels (P=0.016 and 6.1 \times 10^{-3} ) (Table 2) and increased HOMA-IR values (P=0.041) among glucose-tolerant subjects.

Body mass index, HbA1c and circulating leptin levels were not associated with carrier-status of the rs881858. Moreover, we investigated measures of insulin release derived from OGTT data in relation to rs881858, but the insulinogenic index and the disposition index were not different between genotypes (Table 2).
Association of G-allele of VEGFA rs881858 with fasting circulating TSH, thyroid hormone levels and metabolic traits after 5-yr follow-up

We studied changes in metabolic traits over the 5 yr follow-up period were studied among glucose tolerant individuals from the Inter99 cohort (Table S2, Table S6). Glucose (HbA1c: 0.66% and fasting plasma glucose: 0.86%) and insulin measures (fasting serum insulin: 3.2%; HOMA-IR: 4.1%) improved over a mean follow-up time of 5.4 years (Table S2). However, none of these measures were associated with the VEGFA rs881858 polymorphism. Moreover, changes in BMI, circulating TSH and T4 levels were also not associated with VEGFA rs881858 (Table S6). This indicates that the variant could act by modulating a given set point for TSH, since the phenotype of carriers appears to be stable during follow-up.

Reporter-gene analysis of VEGFA rs881858 and rs9472138 alleles

To determine if the most significant TSH associated VEGFA SNPs rs881858 or rs9472138 could confer functional changes to the VEGFA promoter activity, luciferase reporter vectors representing the SNPs were tested by transfection in HEK293 cells followed by luciferase assays (Fig. 2). Basal activity of the VEGFA minimal promoter was very high in HEK293 cells (not shown). The activity of the rs881858 A allele was significantly higher than the G-allele (P=0.0012) (Fig. 2A), while there was no difference between the C and the T-allele of VEGFA rs9472138.

The A-allele of VEGFA rs881858 is predicted to create a novel binding site for the transcription factor CHOP. To test the response of this site to CHOP, HEK293 cells were transfected with increasing amounts of CHOP expression vector in the presence of either rs881858 A- or G-allele reporter vector (Fig. 2B). With no CHOP over-expression the VEGFA A-allele had increased activity compared with the G-allele (as in Fig. 2A). Low amounts of CHOP expression vector increased the G-allele reporter activity, while there was no difference in the activity of the A-allele. Furthermore,
increasing amounts of CHOP activity resulted in significantly decreased activity of the A-allele, indicating that the activity of this site is repressed by CHOP. Thus, the main action of CHOP on the predicted binding site created by rs881858 A was to decrease reporter gene activity, which is in line with CHOP being a transcriptional repressor. When over-expressing c/EBPβ both the VEGFA A- and the G-alleles of rs881858 responded by increasing luciferase activities 25% (P<0.05) at low levels of c/EBPβ while decreasing at higher amounts of c/EBPβ (Fig. 2C). Since the rs881858 is not located in the core binding site of c/EBPβ (Fig. 1B), this is compatible with an equal response to c/EBPβ by either SNP allele.

The interaction between different amounts of c/EBPβ and CHOP was tested by co-transfection experiments in which varying ratios of c/EBPβ and CHOP were used with reporter vectors. For all combinations of CHOP in the presence of c/EBPβ, CHOP repressed the A-allele luciferase activity, while having no effect on the G-allele (Fig. 2D). Thus, based on reporter-gene assays, the A-allele of VEGFA rs881858 creates a novel response element of CHOP effectively repressing the minimal promoter activity of VEGFA, while this has no effect on the response to c/EBPβ. Furthermore, the A-allele confers higher reporter-gene activity in the basal state compared to the G-allele.

**Binding affinities of VEGFA rs881858 (A/G) alleles**

Electrophoretic mobility shift assays (EMSA) were made assess in vitro transcription factor binding of VEGFA rs881858. Double-stranded (ds)-oligos representing the A-allele and the G-allele and encompassing both the c/EBPβ and the CHOP binding sites were compared with known c/EBPβ and the CHOP binding sites. Labeled ‘A’ and ‘G’ oligos formed two strong binding complexes (lane 1 and 6) of which the double band could be removed by competition (COMP) with unlabelled A and G-oligos, as well as with un-labeled CHOP (lane 2 and 9) and c/EBPβ (lane 5) oligos (Fig. 3A). This complex contains CHOP as well as c/EBPβ protein, because of the efficient competition by corresponding unlabelled oligos, and the reduction in band intensity, when adding C/EBPβ antibody
to the binding reaction (lane 10, ‘supershift’). Moreover, the intensity of the CHOP complex was increased for the A oligo compared with G, indicating stronger binding of the A probe to CHOP protein (Fig. 3A, lane 1 vs. lane 6, Fig. 3B, lane 5 vs. 6 and 9 vs. 10). One complex was specific for A and G oligos as these efficiently out-compete the radio-labeled probes (Fig. 3A and 3B, arrow), whereas competition using oligos containing known CHOP or c/EBPβ sites did not remove the complex (Fig. 3A, lane 1 vs. 2, 5 vs. 6, 5 vs. 9 and Fig. 3B, lane 1 vs. 2). Thus, both A and G versions of rs881858 can bind CHOP and c/EBPβ \textit{in vitro}.

We also compared A and G oligos with oligos representing known CHOP or C/EBPβ binding sites\(^{30,31}\) (Fig. 3B). CHOP and C/EBPβ oligos form complexes with the same mobility shift as A and G oligos. The same lower complex formed with either the CHOP or C/EBPβ probe, consistent with CHOP and C/EBPβ forming heterodimers. UPR induction increased C/EBPβ, A and G complex quantity (Fig. 3B). Furthermore, excess unlabelled CHOP oligo efficiently removed the VEGFA A and G probe binding at the lower complex, but not the top complex (arrow, Fig. 3B). The A oligo consistently formed more lower complex than the G oligo, indicating increased binding strength of this probe. The top complex, specific for the rs881858 site (Fig. 3A and 3B, arrows) suggests that additional proteins may bind the \textit{VEGFA} rs881858 site.

The c/EBPβ probe formed a faint complex (lane 3 and 7 Fig. 3B, lanes 1-3, Fig. 3C), most likely c/EBPβ homo-dimer, because it was super-shifted with c/EBPβ antibody and competed with unlabelled c/EBPβ probe. The known CHOP binding site of the \textit{TRIB3} promoter was used as probe for EMSA (Fig. 3C, lane 4-10). As expected the formed complex contained CHOP, shown by efficient competition by unlabelled CHOP ds-oligo (lane 4, Fig. 3C) and decreased complex formation with addition of CHOP antibody (lane 5, Fig. 3C). However, since excess of A or G did not remove TRIB3 probe binding, this indicates that the investigated \textit{VEGFA} binding site has lower affinity for CHOP than the TRIB3 site with no difference between the A-allele and the G-allele of
the rs881858. In summary, the VEGFA rs881858 site forms a binding site for both CHOP and c/EBPβ, where the A-allele forms more CHOP complex compared with the G-allele, but where this site measured against a well-characterized CHOP binding site has lower affinity for CHOP.

Public eQTL databases were searched in order to determine, if base-line thyroid VEGFA mRNA levels correlated with the SNPs significantly associated with TSH. Data from the GTEx database\textsuperscript{32} is shown in Figure S4 indicating that base-line thyroid VEGFA transcript levels do not depend on genotypes of the SNPs rs881858, rs9472138, rs943080 or rs4711751.

DISCUSSION AND CONCLUSIONS

We focused on the genetic and molecular characterization of the VEGFA locus with the aim to identify the functional variant(s) explaining the association with circulating TSH levels\textsuperscript{14-17}. We performed a dense association-mapping for VEGFA SNPs using data from 8445 Danish individuals, and identified 4 SNPs significantly associated with circulating TSH levels, all of which were located more than 50kb 3' of the coding region. Among the top hits we identified VEGFA rs881858, in high LD to the VEGFA GWAS SNP rs9472138\textsuperscript{14}. The carriers of VEGFA rs881858 G allele had a 2.3% decrease in circulating TSH levels in an additive manner. Moreover, the identified SNP rs881858 provided functional evidence of allele specific effects at a VEGFA regulatory region binding CHOP and C/EBPβ proteins, thus connecting cellular stress activated pathways with VEGFA gene regulatory activity and thyroid function, because CHOP production is activated by several types of cellular stress, such as endoplasmic reticulum stress, nutrient deprivation or oxidative stress\textsuperscript{33}.

Thus, data here and elsewhere\textsuperscript{14} support that common variation in the VEGFA locus is an important determinant of circulating TSH levels. For the VEGFA locus, the variants most highly associated with TSH, rs881858 and rs9472138 are located >50kb 3' of the VEGFA coding region, and equally
close to a long non-coding RNA (Loc100132354), the function of which is not characterized. However, chromosom al interaction analysis identified contact points between the rs881858 SNP region and the promoter of VEGFA indicating control of VEGFA gene activity (Fig. 1A).

VEGFA is important for angiogenesis, homeostatic responses and organ growth in multiple tissues or cell types (white adipose tissue\textsuperscript{34,35}, islets of Langerhans\textsuperscript{36,37}), and has also been shown to control T4 to T3 conversion in hypothalamic tanycytes and thereby the feedback control of thyroid hormones to TRH and TSH release\textsuperscript{38,39}. VEGFA rs881858 is also a known GWAS locus for chronic kidney disease (CKD) and kidney function\textsuperscript{40} suggesting a possibility that this SNP controls responses in multiple organs, also as VEGFA is expressed ubiquitously. Since T4 levels were correspondingly increased in G-allele carriers, whose TSH levels were decreased, rs881858 seem to act primarily on the thyroid gland.

The G-allele is associated with increased VEGFA response to TSH, an increased T4 release, and presumably increased T3 conversion in the hypothalamus and therefore resulting in decreased TSH release as a hypothalamic-pituitary response. However, it is also a possibility that the action of rs881858 can be on both VEGFA expression in the thyroid gland as well as on the hypothalamic tanycytes. Consistent with CHOP being a repressor\textsuperscript{41} A-allele carriers have lower T4 and increased TSH: For A-carriers the induction of ER-stress may result in an impaired response to stimulate thyroid growth via VEGFA due to increased CHOP binding (Fig.4). Since thyroid VEGF is increased by TSH\textsuperscript{19}, this is consistent with rs881858 being more important in thyroid than in hypothalamus. Moreover, CHOP deficient mice have increased angiogenesis, showing that CHOP normally acts to limit angiogenesis\textsuperscript{42}.

Interestingly, the genomic region containing rs881858 and rs9472138 is only modestly associated with levels of circulating VEGF, while regions situated both 5’ of the VEGFA coding regions and further 3’ in the C6orf223 locus are highly associated with circulating VEGF levels\textsuperscript{43,44}. There is
little or no LD between these regions and the LD block harbouring the SNPs associated with circulating TSH in the current study (rs881858, rs94772138, rs943080 and rs4711751) and these signals appear to be independent. Thus, it is conceivable that several regions in or near the VEGFA locus control different aspects of VEGF regulation and release. It seems reasonable that regulatory mechanisms operating in thyroid tissue to govern for example stress induced VEGF production could differ from the molecular mechanisms contributing to the regulation of circulating VEGF levels. Although we do not know the contribution of the thyroid gland to the sVEGF pool it is likely to be only a minor contributor compared with endothelial-released VEGF.

Although we present evidence for differential effects of alleles of VEGFA rs881858 by both reporter assays and direct binding, and have retrieved and inspected SNPs with high LD to this allele; we have not performed an extensive investigation of all linked SNPs in the VEGFA region for evidence of regulatory activity, and thus cannot exclude that additional functional SNPs may exist. Another limitation of our study is that we have no available data to show association between rs881858 alleles and levels of VEGFA mRNA transcript, protein levels or TSH stimulated VEGF release. Data from public eQTL databases shows no association between the investigated SNPs and thyroid VEGFA mRNA levels (Fig. S4).

When examining metabolic traits association with VEGFA rs881858 GG homozygous subjects were slightly more insulin resistant, while having similar measures of obesity and similar insulin secretion capacity (Table 2). Adipose tissue-specific knock-out of Vegfa results in inability to expand the adipose tissue, when demands for fat storage increases, exemplified by high-fat feeding. Our results indicate that in the human setting the A-allele of VEGFA rs881858 via CHOP binding generates lower reporter-gene activity suggesting a decreased VEGFA response to cellular stress. This would result in an impaired angiogenic response of the A-allele, which is consistent with increased TSH levels. However, this is seemingly at odds with the GG-homozygous subjects being more insulin
resistant, because in mouse models insulin resistance is observed when the angiogenic response and therefore adipogenesis is impaired. Thus, further studies are necessary to determine the tissue-specific effects of VEGFA and genetic variation on different human tissues. The VEGFA SNP rs9472138 has previously been associated with visceral obesity and insulin resistance in women, further underlining the importance of genetic variation in VEGFA also for human adipose tissue expansion. Of note, increased circulating TSH is associated with obesity and impaired cardio-metabolic health, which could indirectly affect the observed associations between insulin resistance and VEGFA rs881858. Mendelian randomization studies in large population based cohort could potentially resolve this.

ACKNOWLEDGMENTS

This study was supported by the Danish Medical Research Council (grant DFF – 1331-00033), The Novo Nordisk Foundation, the Danish Diabetes Association, the Gerda and Aage Haensch Foundation, the A.P. Moller Foundation for the Advancement of Medical Science, Novo Nordisk A/S, Roskielde University and University of Copenhagen. The Novo Nordisk Foundation Center for Basic Metabolic Research is an independent Research Center at the University of Copenhagen partially funded by an unrestricted donation from the Novo Nordisk Foundation (www.metabol.ku.dk). Further, this work is supported by the European Commission as an integrated project under the 6th Framework Programme (LSHM-CT-2005-018734, HepAdip). The Danish Obesity Research Centre (DanORC; www.danorc.dk) is supported by The Danish Council for Strategic Research (grant 2101-06-0005). TSA was funded by the Gene Diet Interactions in Obesity (GENDINOB, www.gendinob.dk) project and acknowledges the same. The authors wish to thank Christa Persson, Annemette Forman, Tina Lorentzen, and Marianne Stendal for technical assistance, Arne Nielsen for data management, and Grete Lademann and Trine Toldsted for secretarial support.

References


Legends to figures

Figure 1: A) Schematic representation of SNPs in the VEGFA gene associated with circulating TSH, serum VEGF, Type 2 diabetes or insulin resistance. Chromosomal base pair annotations is given for the hg19 assembly. The citations for the articles is: Porcu et al. ¹⁴, Debbette et al. ⁴⁴, Choi et al. ⁴³, Burgdorf et al. ⁴⁵ and Bonnefond et al. ⁴⁶. B) Genomic region surrounding the VEGFA gene. Simplified representations of ENCODE sub-tracks from Genome Browser are shown. TSS: transcription start site, PROM: promoter, H3K27AC: Histone 3, Lysine 27 Acetylation (a mark of active and regulatory genomic DNA), H3K4Me1: Histone 3, Lysine 4 Mono-Methylation (a mark of active enhancers), Conservation: 24 placental mammals, ChIA-PET: Chromatin Interaction Analysis Paired-End Tags (ChIA-PET) from ENCODE/Genome Institute of Singapore-Ruan ⁴⁷ ⁴⁸, c/EBPβ binding; Chromatin-immunoprecipitation using antibody directed against C/EBPβ followed by sequencing, DNase HS: DNase hyper sensitivity. Genome browser screen-shots of the area are shown in Suppl. Fig. S3. B) The binding sites for c/EBPβ and CHOP aligned showing the binding preferences of c/EBPβ and CHOP at the position of rs881858. TSS: Transcription start site. UIPAC nucleotide abbreviations N: Any, M: A or C, R: A or G, K: G or T, D: A or G or T. Binding site logos were from http://motifmap.ics.uci.edu/⁴⁹.

Figure 2: Reporter gene activities of examined VEGFA gene variants: rs881858 A or G, and rs9472138 C or T, respectively, refer to enhancer plasmids containing the VEGFA minimal promoter as well as the DNA regions surrounding rs881858 and rs9472138 and representing the different version of the SNPs. A) Basal activity of VEGFA minimal promoter and SNP-containing regions. Shown are relative luciferase activities of plasmids transfected into HEK203 cells. pGL4.10: promoter-less plasmid. Min pro: VEGFA minimal promoter plasmid. Luciferase activities were normalized to beta-galactosidase activity and are presented relative to the activity of the VEGFA minimal promoter, B) and C) Response of the rs881858 A and G alleles to separate CHOP and c/EBPβ over-expression, respectively. Shown are relative luciferase activities of rs881858 enhancer plasmids harbouring A or G co-transfected with CHOP or c/EBPβ expression plasmids in HEK293 cells. Luciferase activities were normalized to beta-galactosidase activity and are presented relative to the activity of the VEGFA minimal promoter (not shown on graph). D) Response of the rs881858 A and G alleles to varying levels of concurrent CHOP and c/EBPβ over-expression (5ng per well). Relative luciferase activities of rs881858 enhancer plasmids harbouring A or G co-transfected with CHOP and c/EBPβ expression plasmids in HEK293 cells. Experiments were performed 4 times in triplicate. * P<0.05, **P<0.01 by t-test for the indicated comparison or compared against the basal activity of the allele.

Figure 3: Electrophoretic Mobility Shift Assay (EMSA) experiments probing the VEGFA rs881858 region for binding using HEK293 nuclear extract. A) Lane 1-4: ³²P-labelled A-allele as probe (‘A’), 5-10: ³²P-labelled G-allele as probe (‘G’). Nuclear extracts from DTT treated HEK293 in all lanes. B) Comparison of complex formation using probes for the A-allele, G-allele, CHOP and c/EBPβ binding sites. C) Identification of complexes binding to CHOP and C/EBPβ using known binding sites for these proteins. Abbreviations: Probe: ³²P-labelled ds-oligo as indicated. COMP: Competition using unlabelled ds-oligo, Ab.: Antibody used for super-shift of complexes. N.E.: Nuclear Extract. CHOP: ds-oligo having the CHOP binding site from the TRIB3 (tribbles pseudokinase 3) gene ³¹.

https://mc.manuscriptcentral.com/jmedgenet
CEBP: ds-oligo having the c/EBPβ binding site from the c/EBPα gene promoter. Labels to the left of images indicate formed complexes: c/EBPβ homodimer, CHOP/c/EBPβ heterodimer and arrows indicate rs881858 specific complexes not binding CHOP or c/EBPβ. N.s.: non-specific. Shown are representative blots (n=2-4 of each).

**Figure 4:** A model illustrating the possible involvement of *VEGFA* rs881858 in regulating thyroid function and the set point of TSH. The rs881858 A-allele is repressed by CHOP, which is activated by the cellular stress, resulting in less VEGFA enhancer activity, with predicted less thyrocyte VEGF produced. The result is predicted to be reduced angiogenesis and less thyroid compensatory expansion. This result will be less T4 production, less feedback inhibition of TRH and TSH release and therefore increased circulating TSH levels. A higher TSH level will result in increased stimulation of the TSH receptor and a higher degree of activation of the ER-stress response. Dashed lines illustrate rs881858 A-allele related effects.
Supplementary Figure legends

Figure S1: Meta-Analysis Forest plot of the association between TSH and the G-allele of VEGFA rs881858

The association between TSH and the G-allele of rs881858 of the VEGFA locus was meta-analysed in three cohorts: Inter99, Health2006 and Health2008. Shown are the β-values and SE for each cohort with corresponding 95% confidence intervals and overall P-value evaluated using a fixed effect model. A total of 8,445 individuals from three Danish cohorts (Inter99, Health2006, and Health2008) were combined for inverse variance meta-analyses, where weights are proportional to the squared standard errors of the effect estimates. Genomic inflation factor (λ) was at acceptable levels (λTSH = 1.0) after the meta-analyses. A chi-square test for heterogeneity (I) was implemented, to estimate the heterogeneity in effect sizes across different participating cohorts using METAL software (http://csg.sph.umich.edu/abecasis/metal/).

Figure S2. LD Heat Map of pairwise r^2 values of SNPs studied from the VEGFA gene region

LD estimations and proxy search were performed using 1000 genomes project data implemented in SNP Annotation Proxy search tool (http://www.broadinstitute.org/mpg/snap/). An LD heat-map depicting pairwise r^2 values as colors (least correlated r2=0 as dark blue and most correlated r2=1.0 as light shade of blue). Non-genotyped SNPs in LD with TSH associated SNPs were retrieved from the ENSEMBL genome browser (http://grch37.ensembl.org/Homo_sapiens/Info/Index) using CEU data, only SNPs with LD≥0.8 were investigated further for evidence of differential allele effects.

Figure S3: Genome browser (www.genome.ucsc.edu) screen shots of genomic regions containing the SNPs summarized in Table S5.

Figure S4: Expressed quantitative trait locus (eQTL) data for rs9472138, rs881858, rs943080 and rs4711751 for VEGFA mRNA in thyroid tissue. The mRNA levels of VEGFA are presented according to genotype of each of the SNPs. Data were retrieved from the GTEX portal (www.gtexportal.org)\textsuperscript{32}.
Figure 1: A) Schematic representation of SNPs in the VEGFA gene associated with circulating TSH, serum VEGF, Type 2 diabetes or insulin resistance. Chromosomal base pair annotations is given for the hg19 assembly. The citations for the articles is: Porcu et al.14, Debbette et al. 44, Choi et al. 43, Burgdorf et al. 45 and Bonnefond et al. 46. B) Genomic region surrounding the VEGFA gene. Simplified representations of ENCODE sub-tracks from Genome Browser are shown. TSS: transcription start site, PROM: promoter, H3K27AC: Histone 3, Lysine 27 Acetylation (a mark of active and regulatory genomic DNA), H3K4Me1: Histone 3, Lysine 4 Mono-methylation (a mark of active enhancers), Conservation: 24 placental mammals, ChIA-PET: Chromatin Interaction Analysis Paired-End Tags (ChIA-PET) from ENCODE/Genome Institute of Singapore-Ruan47 48, c/EBP-β binding; Chromatin-immunoprecipitation using antibody directed against c/EBPβ followed by sequencing, DNase HS: DNase hyper sensitivity. Genome browser screen-shots of the area are shown in Suppl. Fig. S3. B) The binding sites for c/EBPβ and CHOP aligned showing the binding preferences of c/EBPβ and CHOP at the position of rs881858. TSS: Transcription start site. UIPAC nucleotide abbreviations: N: Any, M: A or C, R: A or G, K: G or T, D: A or G or T. Binding site logos were from
http://motifmap.ics.uci.edu/49.
Fig. 1
209x297mm (300 x 300 DPI)
Figure 2: Reporter gene activities of examined VEGFA gene variants: rs881858 A or G, and rs9472138 C or T, respectively, refer to enhancer plasmids containing the VEGFA minimal promoter as well as the DNA regions surrounding rs881858 and rs9472138 and representing the different version of the SNPs. A) Basal activity of VEGFA minimal promoter and SNP-containing regions. Shown are relative luciferase activities of plasmids transfected into HEK203 cells. pGL4.10: promoter-less plasmid. Min pro: VEGFA minimal promoter plasmid. Luciferase activities were normalized to beta-galactosidase activity and are presented relative to the activity of the VEGFA minimal promoter. B) and C) Response of the rs881858 A and G alleles to separate CHOP and c/EBPβ over-expression, respectively. Shown are relative luciferase activities of rs881858 enhancer plasmids harbouring A or G co-transfected with CHOP or c/EBPβ expression plasmids in HEK293 cells. Luciferase activities were normalized to beta-galactosidase activity and are presented relative to the activity of the VEGFA minimal promoter (not shown on graph). D) Response of the rs881858 A and G alleles to varying levels of concurrent CHOP and c/EBPβ over-expression (5ng per well). Relative luciferase activities of rs881858 enhancer plasmids harbouring A or G co-transfected with CHOP and c/EBPβ.
expression plasmids in HEK293 cells. Experiments were performed 4 times in triplicate. * P<0.05, **P<0.01 by t-test for the indicated comparison or compared against the basal activity of the allele.

Fig. 2
197x489mm (600 x 600 DPI)
Figure 3: Electrophoretic Mobility Shift Assay (EMSA) experiments probing the VEGFA rs881858 region for binding using HEK293 nuclear extract. A) Lane 1-4: 32P-labeled A-allele as probe (‘A’), 5-10: 32P-labeled G-allele as probe (‘G’). Nuclear extracts from DTT treated HEK293 in all lanes. B) Comparison of complex formation using probes for the A-allele, G-allele, CHOP and c/EBPβ binding sites. C) Identification of complexes binding to CHOP and C/EBPβ using known binding sites for these proteins. Abbreviations: Probe: 32P-labeled ds-oligo as indicated. COMP: Competition using unlabelled ds-oligo, Ab.: Antibody used for super-shift of complexes. N.E.: Nuclear Extract. CHOP: ds-oligo having the CHOP binding site from the TRIB3 (tribbles pseudokinase 3) gene 31. CEBP: ds-oligo having the c/EBPβ binding site from the c/EBPα gene promoter 30. Labels to the left of images indicate formed complexes: c/EBPβ homodimer, CHOP/c/EBPβ heterodimer and arrows indicate rs881858 specific complexes not binding CHOP or c/EBPβ. N.s.: non-specific. Shown are representative blots (n=2-4 of each).

Fig. 3
209x297mm (300 x 300 DPI)
Figure 4: A model illustrating the possible involvement of VEGFA rs881858 in regulating thyroid function and the set point of TSH. The rs881858 A-allele is repressed by CHOP, which is activated by the cellular stress, resulting in less VEGFA enhancer activity, with predicted less thyrocyte VEGF produced. The result is predicted to be reduced angiogenesis and less thyroid compensatory expansion. This result will be less T4 production, less feedback inhibition of TRH and TSH release and therefore increased circulating TSH levels. A higher TSH level will result in increased stimulation of the TSH receptor and a higher degree of activation of the ER-stress response. Dashed lines illustrate rs881858 A-allele related effects.
Figure S2

rs881858

<table>
<thead>
<tr>
<th>Study</th>
<th>Beta</th>
<th>SE</th>
<th>95%-CI W(fixed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inter99</td>
<td>-0.046</td>
<td>0.009</td>
<td>[-0.064; -0.028]</td>
</tr>
<tr>
<td>Health2006</td>
<td>-0.047</td>
<td>0.015</td>
<td>[-0.076; -0.018]</td>
</tr>
<tr>
<td>Health2008</td>
<td>-0.038</td>
<td>0.028</td>
<td>[-0.093; 0.017]</td>
</tr>
</tbody>
</table>

Fixed effect model

Heterogeneity: I-squared=0%, p=0.9588

-0.05 0 0.05

-0.046 [-0.060; -0.031] 100%
<table>
<thead>
<tr>
<th>Chromosome</th>
<th>Start Base</th>
<th>End Base</th>
<th>Gene/Region</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>100</td>
<td>200</td>
<td>cMyc</td>
</tr>
<tr>
<td>2</td>
<td>500</td>
<td>600</td>
<td>Pol2</td>
</tr>
<tr>
<td>3</td>
<td>2000</td>
<td>4000</td>
<td>TBP</td>
</tr>
<tr>
<td>4</td>
<td>5000</td>
<td>7000</td>
<td>GABP</td>
</tr>
</tbody>
</table>

**Gene Expression**

- **cMyc**: High expression in K562 and GM12878
- **Pol2**: Low expression in H1-hESC
- **TBP**: Moderate expression in GM12878
- **GABP**: High expression in K562

**ChIP-seq Analysis**

- **K562 cMyc**: ChIP-seq信号重叠
- **H1-hESC cMyc**: ChIP-seq信号重叠
- **GM12878 Pol2**: ChIP-seq信号重叠

**Repeat Masking**

- **RepeatMasker**: Repeat elements
- **Multiz Alignments**: 100 Vertebrates
MCF-7 (serum_stimulated) TFBS Uniform Peaks of c-Myc from ENCODE/UT-A/Analysis
MCF-7 (serum_stimulated) TFBS Uniform Peaks of Pol2 from ENCODE/UT-A/Analysis
H1-hESC TFBS Uniform Peaks of Pol2-4H8 from ENCODE/HudsonAlpha/Analysis
HUVEC TFBS Uniform Peaks of Pol2-4H8 from ENCODE/HudsonAlpha/Analysis
HepG2 TFBS Uniform Peaks of Pol2-4H8 from ENCODE/HudsonAlpha/Analysis
GM12878 TFBS Uniform Peaks of SP1 from ENCODE/HudsonAlpha/Analysis
GM12878 Pol2-4H8 PCR1x ChIP-seq Raw Signal Rep 2 from ENCODE/HAIB
DNaseI/FAIRE/ChIP Synthesis from ENCODE/OpenChrom(Duke/UNC/UTA)
K562 Pol2-4H8 v041610.1 ChIP-seq Raw Signal Rep 2 from ENCODE/HAIB
H1-hESC CTCF TFBS ChIP-seq Density Signal ENCODE/OpenChrom-UTA
GM12878 Pol2 TFBS ChIP-seq Overlap Signal ENCODE/OpenChrom-UTA
H1-hESC Pol2 TFBS ChIP-seq Overlap Signal ENCODE/OpenChrom-UTA
GM12878 USF-1 PCR2x ChIP-seq Raw Signal Rep 2 from ENCODE/HAIB
GM12878 GABP PCR2x ChIP-seq Raw Signal Rep 1 from ENCODE/HAIB
GM12878 CTCF TFBS ChIP-seq Peaks from ENCODE/OpenChrom-UTA
K562 GABP v041610.1 ChIP-seq Raw Signal Rep 1 from ENCODE/HAIB
HeLa-S3 TFBS Uniform Peaks of HA-E2F1 from ENCODE/USC/Analysis
K562 CTCF TFBS ChIP-seq Overlap Signal ENCODE/OpenChrom-UTA
H1-hESC Pol2 TFBS ChIP-seq Peaks from ENCODE/OpenChrom-UTA
K562 Pol2 TFBS ChIP-seq Overlap Signal ENCODE/OpenChrom-UTA
Monocytes-CD14+ DNaseI HS Uniform Peaks from ENCODE/Analysis
K562 Pol2 TFBS ChIP-seq Density Signal ENCODE/OpenChrom-UTA
GM12878 USF-1 PCR2x ChIP-seq Peaks Rep 1 from ENCODE/HAIB
GM12878 USF-1 PCR2x ChIP-seq Peaks Rep 2 from ENCODE/HAIB
H1-hESC USF-1 PCR1x ChIP-seq Peaks Rep 2 from ENCODE/HAIB
H1-hESC GABP PCR1x ChIP-seq Peaks Rep 2 from ENCODE/HAIB
K562 USF-1 v041610.1 ChIP-seq Peaks Rep 2 from ENCODE/HAIB
K562 GABP v041610.1 ChIP-seq Peaks Rep 2 from ENCODE/HAIB
K562 cMyc TFBS ChIP-seq Peaks from ENCODE/OpenChrom-UTA
GM12878 CHD2 IgG-mus ChIP-seq Signal from ENCODE/SYDH
GM12878 CHD2 IgG-mus ChIP-seq Peaks from ENCODE/SYDH
Histone Modifications by ChIP-seq from ENCODE/Broad Institute
NHGRI Catalog of Published Genome-Wide Association Studies
GM12878 Pol2 Standard ChIP-seq Peaks from ENCODE/SYDH
GM12878 Pol2 IgG-mus ChIP-seq Signal from ENCODE/SYDH
K562 ZNF274 UC Davis ChIP-seq Signal from ENCODE/SYDH
K562 HDAC8 GFP-tag TFBS Signal from ENCODE/UChicago
K562 NR4A1 GFP-tag TFBS Signal from ENCODE/UChicago
K562 NR4A1 GFP-tag TFBS Peaks from ENCODE/UChicago
K562 JunB GFP-tag TFBS Signal from ENCODE/UChicago
K562 FOS GFP-tag TFBS Peaks from ENCODE/UChicago
X. tropicalis (Nov. 2009 (JGI 4.2/xenTro3)) Alignment Net
Opossum (Oct. 2006 (Broad/monDom5)) Alignment Net
Zebrafish (Jul. 2010 (Zv9/danRer7)) Alignment Net
Repeating Elements by RepeatMasker
rs9394969
rs9394969
Supplementary Figure S4: Thyroid eQTL data (GTEX portal.org)

<table>
<thead>
<tr>
<th>SNP</th>
<th>P-Value</th>
<th>Effect Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs9472138</td>
<td>0.15</td>
<td>0.067</td>
</tr>
<tr>
<td>rs881858</td>
<td>0.22</td>
<td>-0.058</td>
</tr>
<tr>
<td>rs943080</td>
<td>0.24</td>
<td>0.049</td>
</tr>
<tr>
<td>rs4711751</td>
<td>0.26</td>
<td>-0.047</td>
</tr>
</tbody>
</table>
Supplementary tables

Table S1: Baseline characteristics of the three Danish cohorts participating in discovery analyses.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Inter99</th>
<th>Health2006</th>
<th>Health2008</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>5645</td>
<td>2711</td>
<td>601</td>
</tr>
<tr>
<td>Men (%)</td>
<td>50.1</td>
<td>44.8</td>
<td>43.7</td>
</tr>
<tr>
<td>Age (yrs)</td>
<td>46.1 ± 7.9</td>
<td>48.9 ± 13.1</td>
<td>46.4 ± 8.1</td>
</tr>
<tr>
<td>BMI (kg/m²)</td>
<td>26.2 ± 4.5</td>
<td>25.8 ± 4.6</td>
<td>25.6 ± 4.3</td>
</tr>
<tr>
<td>TSH (mIU/L)</td>
<td>1.39 ± 0.67</td>
<td>1.53 ± 0.73</td>
<td>1.15 ± 0.55</td>
</tr>
</tbody>
</table>

Data are mean ± standard deviation.

Table S2: Study population characteristics at baseline and at five years follow-up in the Inter99 normal glucose tolerant (NGT) participants.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Baseline</th>
<th>Follow up</th>
<th>Changes from baseline to follow-up</th>
</tr>
</thead>
<tbody>
<tr>
<td>n (%men)</td>
<td>4,374 (46.3%)</td>
<td>Up to 3,467 (46.4%)</td>
<td></td>
</tr>
<tr>
<td>Age (years)</td>
<td>45.1 ± 7.8</td>
<td>50.5 ± 7.8</td>
<td>5.38 ± 2.3</td>
</tr>
<tr>
<td>BMI (kg/m²)</td>
<td>25.5 ± 4.0</td>
<td>25.7 ± 4.0</td>
<td>0.44 ± 1.7</td>
</tr>
<tr>
<td>HbA1c (%)</td>
<td>5.7 ± 0.3</td>
<td>5.7 ± 0.3</td>
<td>-0.038 ± 0.26</td>
</tr>
<tr>
<td>Leptin (ng/ml)</td>
<td>5.3 (2.4-11.1)</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>T4 (pmol/L)</td>
<td>15 (13.9-16.3)</td>
<td>14.7 (13.6-15.9)</td>
<td>-0.3 (-1.3-0.6)</td>
</tr>
<tr>
<td>TSH (mU/L)</td>
<td>1.26 (0.91-1.75)</td>
<td>1.41 (1.0-1.99)</td>
<td>0.14 (-0.15-0.5)</td>
</tr>
<tr>
<td>Fasting plasma glucose (mmol/L)</td>
<td>5.3 ± 0.4</td>
<td>5.2 ± 0.5</td>
<td>-0.046 ± 0.49</td>
</tr>
<tr>
<td>2hr glucose (mmol/L)</td>
<td>5.4 ± 1.1</td>
<td>5.5 ± 1.4</td>
<td>0.016 ± 1.48</td>
</tr>
<tr>
<td>Fasting serum insulin (pmol/L)</td>
<td>31.0 (22.0 - 45.0)</td>
<td>28.0 (21.0 - 41.0)</td>
<td>-1.0 (-11.0-8.0)</td>
</tr>
<tr>
<td>Insulin sensitivity index (ISI_MATSUDA)</td>
<td>3.16 (2.24-4.44)</td>
<td>3.12 (2.18-4.40)</td>
<td>-</td>
</tr>
<tr>
<td>Insulinogenic index</td>
<td>77.1 (49.1-128.0)</td>
<td>77.1 (50.0-124.2)</td>
<td>-</td>
</tr>
<tr>
<td>HOMA-IR</td>
<td>1.21 (0.83-1.80)</td>
<td>1.09 (0.77-1.65)</td>
<td>-0.05 (-0.46-0.34)</td>
</tr>
<tr>
<td>Disposition index</td>
<td>229.2 (161.7-348.2)</td>
<td>230.8 (161.6-337.2)</td>
<td>-</td>
</tr>
</tbody>
</table>

We use interquartile range for insulin and related measures as the distribution may be skewed. Data are mean ± SD or median (interquartile range).
Table S3: Oligonucleotides used for cloning VEGFA promoter-enhancer constructs and electrophoretic mobility shift assays

<table>
<thead>
<tr>
<th>Name</th>
<th>Sequence (italics: tail for fill-in labeling, bold CHOP10 site, red: rs881858)</th>
<th>Purpose</th>
</tr>
</thead>
<tbody>
<tr>
<td>RS881858 G sense</td>
<td>agctTGCTGTTCATGCAATGATCCCGGC</td>
<td>Fill-in labeled probe</td>
</tr>
<tr>
<td>RS881858 G a-sense</td>
<td>agctGCGGATCATTTGCAAACAGCA</td>
<td>Fill-in labeled probe</td>
</tr>
<tr>
<td>RS881858 A sense</td>
<td>agctTGCTGTTCATGCAATGATCCCGGC</td>
<td>Fill-in labeled probe</td>
</tr>
<tr>
<td>c/EBPβ sense</td>
<td>agctGCGGATCATTTGCAAACAGCA</td>
<td>Fill-in labeled probe</td>
</tr>
<tr>
<td>c/EBPβ a-sense</td>
<td>agctTGCTGTTCATGCAATGATCCCGGC</td>
<td>Full length probe for competition assays</td>
</tr>
<tr>
<td>RS881858 G sense f</td>
<td>agctTGCTGTTCATGCAATGATCCCGGC</td>
<td>Full length probe for competition assays</td>
</tr>
<tr>
<td>RS881858 G a-sense f</td>
<td>agctGCGGATCATTTGCAAACAGCA</td>
<td>Full length probe for competition assays</td>
</tr>
<tr>
<td>RS881858 A sense f</td>
<td>agctTGCTGTTCATGCAATGATCCCGGC</td>
<td>Full length probe for competition assays</td>
</tr>
<tr>
<td>RS881858 A sense f</td>
<td>agctGCGGATCATTTGCAAACAGCA</td>
<td>Full length probe for competition assays</td>
</tr>
<tr>
<td>c/EBPβ sense f</td>
<td>agctGCGGATCATTTGCAAACAGCA</td>
<td>Full length probe for competition assays</td>
</tr>
<tr>
<td>CHOP10 sense</td>
<td>agctTGCTGTTCATGCAATGATCCCGGC</td>
<td>Fill-in labeled probe</td>
</tr>
<tr>
<td>CHOP10 a-sense</td>
<td>agctTGCTGTTCATGCAATGATCCCGGC</td>
<td>Fill-in labeled probe</td>
</tr>
<tr>
<td>CHOP10 sense f</td>
<td>agctTGCTGTTCATGCAATGATCCCGGC</td>
<td>Full length probe for competition assays</td>
</tr>
<tr>
<td>CHOP10 a-sense f</td>
<td>agctTGCTGTTCATGCAATGATCCCGGC</td>
<td>Full length probe for competition assays</td>
</tr>
<tr>
<td>VEGFA pro F</td>
<td>CTAACTGGGCCTACCTCTCCCCAAAGGACCCAGTC</td>
<td>Cloning: PCR primer VEGFA promoter-exon 1, sense</td>
</tr>
<tr>
<td>VEGFA pro R</td>
<td>TATCCTGGGCCTAGGCCCTAGGCCTAGGCAATGGGAAT</td>
<td>Cloning: PCR primer VEGFA promoter-exon 1, a-sense</td>
</tr>
<tr>
<td>RS881858 sense</td>
<td>AAAATCGATAAAGGTTCCTCCCTCAGAGAGGACGTCAGAGTGC</td>
<td>Cloning: PCR primer RS881858 region, sense</td>
</tr>
<tr>
<td>RS881858 a-sense</td>
<td>AAAATCGATAAAGGTTCCTCCCTCAGAGAGGACGTCAGAGTGC</td>
<td>Cloning: PCR primer RS881858 region, a-sense</td>
</tr>
<tr>
<td>RS9472138 sense</td>
<td>AAAATCGATAAAGGTTCCTCCCTCAGAGAGGACGTCAGAGTGC</td>
<td>Cloning: PCR primer RS9472138 region, sense</td>
</tr>
<tr>
<td>RS9472138 a-sense</td>
<td>AAAATCGATAAAGGTTCCTCCCTCAGAGAGGACGTCAGAGTGC</td>
<td>Cloning: PCR primer RS9472138 region, a-sense</td>
</tr>
</tbody>
</table>
Table S4: *VEGFA* SNPs not reaching study-wide significance for association with circulating levels of thyroid stimulating hormone (TSH)

<table>
<thead>
<tr>
<th>SNP name</th>
<th>Position (build 37/hg19)</th>
<th>Location wrt <em>VEGFA</em></th>
<th>Alleles (effect/other)</th>
<th>EAF</th>
<th>Inter99 n=5,420</th>
<th>Health2006 n=2,442</th>
<th>Health2008 n=583</th>
<th>Combined</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Effect</td>
<td>P</td>
<td>Effect</td>
<td>Effect</td>
<td>N</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Inter99</td>
<td>Health2006</td>
<td>Health2008</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>I² (P_{HET})</td>
</tr>
<tr>
<td>rs114656313</td>
<td>43,692,999</td>
<td>Upstream</td>
<td>A/C</td>
<td>0.026</td>
<td>-0.012</td>
<td>0.03</td>
<td>0.0007</td>
<td>8,445</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.65</td>
<td>0.45</td>
<td>0.99</td>
<td>0.05</td>
<td>(0.69)</td>
</tr>
<tr>
<td>rs76074477</td>
<td>43,713,214</td>
<td>Upstream</td>
<td>A/G</td>
<td>0.043</td>
<td>0.056</td>
<td>0.0098</td>
<td>-0.047</td>
<td>8,445</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.63</td>
<td>-0.081</td>
<td>0.21</td>
<td>0.45</td>
<td>(0.05)</td>
</tr>
<tr>
<td>rs36208384</td>
<td>43,737,909</td>
<td>5' region</td>
<td>A/C</td>
<td>0.016</td>
<td>0.018</td>
<td>0.56</td>
<td>0.003</td>
<td>8,440</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.57</td>
<td>-0.087</td>
<td>0.47</td>
<td>0.45</td>
<td>(0.67)</td>
</tr>
<tr>
<td>rs74500696</td>
<td>43,748,845</td>
<td>Intron</td>
<td>A/G</td>
<td>0.012</td>
<td>-0.037</td>
<td>0.36</td>
<td>-0.026</td>
<td>8,445</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.95</td>
<td>-0.008</td>
<td>0.94</td>
<td>0.49</td>
<td>(0.83)</td>
</tr>
<tr>
<td>rs998584</td>
<td>43,757,896</td>
<td>Downstream</td>
<td>G/T</td>
<td>0.45</td>
<td>-0.004</td>
<td>0.62</td>
<td>0.016</td>
<td>8,426</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.86</td>
<td>-0.007</td>
<td>0.78</td>
<td>0.57</td>
<td>(0.99)</td>
</tr>
<tr>
<td>rs6905288</td>
<td>43,758,873</td>
<td>Downstream</td>
<td>A/G</td>
<td>0.55</td>
<td>-0.004</td>
<td>0.62</td>
<td>-0.016</td>
<td>8,444</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.25</td>
<td>-0.008</td>
<td>0.74</td>
<td>0.27</td>
<td>(0.79)</td>
</tr>
<tr>
<td>rs68016381</td>
<td>43,761,645</td>
<td>Downstream</td>
<td>T/C</td>
<td>0.047</td>
<td>-0.048</td>
<td>0.021</td>
<td>-0.013</td>
<td>8,445</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.12</td>
<td>-0.050</td>
<td>0.46</td>
<td>0.04</td>
<td>(1.0)</td>
</tr>
<tr>
<td>rs35349911</td>
<td>43,785,255</td>
<td>Downstream</td>
<td>T/C</td>
<td>0.43</td>
<td>-0.001</td>
<td>0.90</td>
<td>0.028</td>
<td>8,445</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.23</td>
<td>-0.013</td>
<td>0.62</td>
<td>0.54</td>
<td>(0.52)</td>
</tr>
<tr>
<td>rs943072</td>
<td>43,795,968</td>
<td>Downstream</td>
<td>A/C</td>
<td>0.09</td>
<td>-0.003</td>
<td>0.82</td>
<td>0.003</td>
<td>8,445</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.88</td>
<td>-0.016</td>
<td>0.73</td>
<td>0.86</td>
<td>(0.92)</td>
</tr>
<tr>
<td>rs145023524</td>
<td>43,819,046</td>
<td>Downstream</td>
<td>A/G</td>
<td>0.006</td>
<td>-0.006</td>
<td>0.91</td>
<td>-0.002</td>
<td>8,445</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.55</td>
<td>0.12</td>
<td>0.40</td>
<td>0.62</td>
<td>(0.66)</td>
</tr>
<tr>
<td>rs55663434</td>
<td>43,820,609</td>
<td>Downstream</td>
<td>A/G</td>
<td>0.015</td>
<td>-0.031</td>
<td>0.39</td>
<td>-0.015</td>
<td>8,445</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.59</td>
<td>0.034</td>
<td>0.73</td>
<td>0.38</td>
<td>(0.83)</td>
</tr>
</tbody>
</table>

*SNPs in LD (r²>0.4). EAF: Effect allele frequency. I²: heterogeneity at meta-analyses level. P_{HET}: P value for heterogeneity.*
Table S5: Evaluation of possible regulatory variants in the \textit{VEGFA} locus.

<table>
<thead>
<tr>
<th>SNP</th>
<th>Position (hg19/chr 6)</th>
<th>Open chromatin (DNase seq)</th>
<th>Conservation</th>
<th>H3K27 acetylation marks</th>
<th>H3K4Me1 marks</th>
<th>Chip-Seq signal</th>
<th>SNP located in TFX BS</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs729761 G&gt;T</td>
<td>43804571</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>+ (RCOR1, NR2F2, TEAD4, GATA2, TAL1)</td>
<td>-</td>
</tr>
<tr>
<td>rs2396083 G&gt;C</td>
<td>43804808</td>
<td>(+)</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>rs2396084 G&gt;A</td>
<td>43804825</td>
<td>(+)</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>rs744103 A&gt;T</td>
<td>43805362</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+ (CMYC)</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>rs10223666 C&gt;G</td>
<td>43805502</td>
<td>(+)</td>
<td>-</td>
<td>-</td>
<td>(+)</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>rs1317983* C&gt;T</td>
<td>43806335</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+ (RAD21, CTCF)</td>
<td>-</td>
</tr>
<tr>
<td>rs881858 G&gt;A</td>
<td>43806609</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+ (CEBPB)</td>
<td>+ (CEBPB &amp; CHOP)</td>
</tr>
<tr>
<td>rs9472135 T&gt;C</td>
<td>43809802</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>rs9472137 T&gt;C</td>
<td>43810469</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>rs9369425 G&gt;A</td>
<td>43810974</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+ (MAFF)</td>
<td>-</td>
</tr>
<tr>
<td>rs9369427 A&gt;C</td>
<td>43811430</td>
<td>-</td>
<td>-</td>
<td>(+)</td>
<td>+</td>
<td>+ (POLR2R, GATA2)</td>
<td>-</td>
</tr>
<tr>
<td>rs9472138 C&gt;T</td>
<td>43811762</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>+ (MAFK, CEBPB)</td>
<td>-</td>
</tr>
<tr>
<td>rs1536304 T&gt;C</td>
<td>43817837</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>+ (FOXA1)</td>
<td>-#</td>
</tr>
<tr>
<td>rs7758685 G&gt;A</td>
<td>43825266</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>rs9394969 G&gt;T</td>
<td>43825459</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>rs943080 T&gt;C</td>
<td>43826627</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>rs4711751 T&gt;C</td>
<td>43828582</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

*rs1317983 is located just 274 nt 5’ of rs881858 and these SNPs share peak for H3K27Ac and H3K4Me1 marks. #FOXA1 site is 2 nt 3’ of SNP, but position is not important for binding. TFX BS: Transcription factor binding site. This table was compiled based on data in Fig. S3.
Table S6: Changes from baseline to follow-up (5yr) in the Inter99 cohort among normal glucose tolerant subjects at baseline for GG, GA and AA genotypes of \textit{VEGFA} rs881858

<table>
<thead>
<tr>
<th>\textit{VEGFA} rs881858</th>
<th>GG</th>
<th>GA</th>
<th>AA</th>
<th>N total</th>
<th>\textit{*Effect}</th>
<th>SE</th>
<th>\textit{P}</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\Delta) TSH (mIU/L)</td>
<td>0.16 (-0.08, 0.47)</td>
<td>0.12 (-0.17, 0.46)</td>
<td>0.14 (-0.17, 0.53)</td>
<td>2843</td>
<td>-0.024</td>
<td>0.088</td>
<td>0.78</td>
</tr>
<tr>
<td>(\Delta) Free T4 (pmol/L)</td>
<td>-0.5 (-1.4, 0.4)</td>
<td>-0.3 (-1.2, 0.8)</td>
<td>-0.3 (-1.3, 0.6)</td>
<td>2856</td>
<td>-0.027</td>
<td>0.067</td>
<td>0.68</td>
</tr>
<tr>
<td>(\Delta) BMI (kg/m(^2))</td>
<td>0.61 ± 1.6</td>
<td>0.41 ± 1.8</td>
<td>0.43 ± 1.6</td>
<td>3466</td>
<td>0.033</td>
<td>0.046</td>
<td>0.47</td>
</tr>
<tr>
<td>(\Delta) HbA1c (%)</td>
<td>-0.05 (-0.17, -0.13)</td>
<td>-0.05 (-0.16, -0.13)</td>
<td>-0.06 (-0.17, -0.13)</td>
<td>2991</td>
<td>0.002</td>
<td>0.007</td>
<td>0.69</td>
</tr>
<tr>
<td>(\Delta) Fasting plasma glucose (mmol/L)</td>
<td>-0.1 (-0.3, 0.2)</td>
<td>-0.1 (-0.3, 0.2)</td>
<td>-0.1 (-0.3, 0.2)</td>
<td>2985</td>
<td>0.005</td>
<td>0.014</td>
<td>0.69</td>
</tr>
<tr>
<td>(\Delta) 2-hour glucose during OGTT (mmol/L)</td>
<td>-0.1 (-0.8, 0.7)</td>
<td>0 (-0.9, 0.8)</td>
<td>-0.1 (-1.0, 0.8)</td>
<td>2965</td>
<td>0.072</td>
<td>0.042</td>
<td>0.091</td>
</tr>
<tr>
<td>(\Delta) Fasting serum insulin (pmol/L)</td>
<td>-2.0 (-12.7, 8)</td>
<td>-1.0 (-11, 9)</td>
<td>-1.0 (-11, 8)</td>
<td>2981</td>
<td>-0.72</td>
<td>0.68</td>
<td>0.28</td>
</tr>
<tr>
<td>(\Delta) HOMA-IR</td>
<td>-0.09 (-0.50, -0.31)</td>
<td>-0.05 (-0.46, -0.34)</td>
<td>-0.04 (-0.46, -0.33)</td>
<td>2979</td>
<td>-0.025</td>
<td>0.028</td>
<td>0.36</td>
</tr>
</tbody>
</table>

* G allele of \textit{VEGFA} as the effect allele assuming an additive genetic model. N total is the number of normal glucose tolerant subjects at baseline who had genotype and phenotype information available during follow-up as well. Values correspond to median (interquartile range) in non-transformed traits.
Table S7. Formulas used for calculating insulinogenic index, the Matsuda insulin sensitivity index, the disposition index and HOMA-IR

<table>
<thead>
<tr>
<th>Trait</th>
<th>Measurement or calculation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Insulinogenic index</td>
<td>(Serum insulin at 30-min (pmol/l) - fasting serum insulin (pmol/l)) / (plasma glucose at 30-min (mmol/l) - fasting plasma glucose (mmol/l))</td>
</tr>
<tr>
<td>( \text{ISI}_{\text{Matsuda}} )</td>
<td>((10,000/\sqrt{\text{fasting plasma glucose (mmol/l)} \times 18 \times \text{fasting serum insulin (pmol/l)} / 6 \times \text{mean plasma glucose (mmol/l)} \times 18 \times \text{mean serum insulin (pmol/l)} / 6 \text{ during OGTT}}))</td>
</tr>
<tr>
<td>Disposition index</td>
<td>The insulinogenic index ( \times ) ( \text{ISI}_{\text{Matsuda}} )</td>
</tr>
<tr>
<td>HOMA-IR</td>
<td>(((\text{fasting serum insulin (pmol/l)} / 6) \times (\text{fasting plasma glucose (mmol/l)})) / 22.5)</td>
</tr>
</tbody>
</table>