Polymorphisms rs55710213 and rs56334587 regulate SCD1 expression by modulating HNF4A binding

Gang Pan *, Marco Cavalli, Claes Wadelius **

Science for Life Laboratory, Department of Immunology, Genetics and Pathology, Uppsala University, Uppsala, Sweden

** Corresponding authors.
E-mail addresses: gang.pan@igp.uu.se (G. Pan), claes.wadelius@igp.uu.se (C. Wadelius).

https://doi.org/10.1016/j.bbagrm.2021.194724
Received 23 February 2021; Received in revised form 16 June 2021; Accepted 17 June 2021
Available online 22 June 2021

1. Introduction

Stearoyl-CoA desaturase 1 (SCD1), encoded by the SCD1 gene at 10q24.31, is the rate limiting enzyme that catalyzes the biosynthesis of MUFAs from SFAs that are either synthesized de novo or derived from diet [1,2]. SCD1 introduces a cis double bond in the Δ9 position of fatty acyl-CoA substrates, primarily palmitoyl-CoA (C16:0) and stearoyl-CoA (C18:0), which are then converted into palmitoleoyl-CoA (C16:1) and oleoyl-CoA (C18:1), respectively [1–3]. The SCD1 activity indices, i.e. the product to precursor fatty acid ratio between C16:1 and C16:0 (desaturation index n-7) and the ratio between C18:1 and C18:0 (desaturation index n-9), have been employed as the surrogate markers for SCD1 activity [4]. SCD1 is believed to play a central role in directing the flow of lipid substrates between energy storage and energy utilization in metabolism [5]. Increased SCD1 expression/activity prompted the biosynthesis of MUFAs, which tend to be predominantly stored as fat in tissues, whereas the precursor SFAs are mainly utilized through fatty acid oxidation [5]. Epidemiological studies, in combination with experimental data, proved that dysregulated SCD1 measured by either the SCD1 activity indices or its expression, was associated with increased risks of many diseases including NAFLD, obesity, hypertension, familial combined hyperlipidemia (FCHL), metabolic syndrome and several types of cancer [2,4,6–11]. In light of its key role in energy metabolism, SCD1 has become a novel therapeutic target for treatment of metabolic disorders and cancers with many small molecule inhibitors targeting SCD1 being developed [5,12]. However, the currently available inhibitors targeting SCD1 have only shown modest effects in inhibiting tumor progression, probably due to the activation of the alternative fatty acid desaturation pathway catalyzed by fatty acid desaturase 2 (FADS2) in many cancers [13,14].

The expression of SCD1 is tightly regulated at transcription level in response to various hormonal and nutrient factors [10]. A variety of transcription factors (TFs), including SREBF1, LXR, PPARα, CEBPA, NF-1, NF-Y, AP-1, SP1, TR and PGC1α, have been identified to regulate the expression of SCD1 by binding to its promoter region [10,12]. Hepatocyte nuclear factor 4 alpha (HNF4A) is a member of the nuclear receptor superfamily of ligand-dependent TFs that is highly expressed in liver and gastrointestinal organs [15,16]. It has been proposed as a central
regulator of glucose and lipid metabolism [16,17]. This is highlighted by inactivating mutations in HNF4A that resulted in maturity-onset diabetes of the young (MODY1) and by the associations of common single nucleotide polymorphisms (SNPs) in the HNF4A locus with the risk of type 2 diabetes in genome-wide association studies (GWAS) and FCHL in family-based linkage analysis [15,18]. The following study suggested that common variants in the HNF4A locus increased the risk of FCHL by upregulating the SCD1 activity indices in FCHL-affected individuals [8]. Accordingly, HNF4A was observed to bind to the promoter region of SCD1 in human hepatocytes [19]. These observations strongly support the hypothesis that SCD1 might be an important downstream target gene that is directly regulated by HNF4A.

Unravelling the regulatory relationship between SCD1 and HNF4A may be helpful in deepening our understanding on the etiologies of various diseases associated with dysregulated SCD1, which may further enable identifying new drug-target mechanisms [5,12]. In this study, we report that HNF4A regulates the expression of SCD1 by binding to the key regulatory regions in the SCD1 locus. The enhancers located both upstream (enhancer E1) and in intron 5 (enhancer E2) of SCD1 are the most important regulatory regions in regulating SCD1 expression. Two adjacent SNPs rs55710213 and rs56334587 in enhancer E2 are functional by being located in a canonical HNF4A binding site. The GG haplotype of these two SNPs disrupted the HNF4A binding site, which significantly decreased the enhancer activity of E2 in response to HNF4A overexpression and is further associated with decreased SCD1 activity.

2. Materials and methods

2.1. Cell culture

HepG2 cells were originally purchased from the American Type Culture Collection (ATCC) and maintained in RPMI1640 basal medium supplemented with 10% fetal bovine serum (FBS) and 2 mM L-glutamine.

Human PLC/PRF/5 cells were purchased from European Collection of Authenticated Cell Cultures (ECACC 85061113) and maintained in high glucose DMEM medium supplemented with 10% FBS, 1 mM sodium pyruvate. Both of the cells were supplemented with 100 units of penicillin and 100 μg of streptomycin per 1 ml of culture medium. The expression of HNF4A in HepG2 and PLC/PRF/5 cells was stably knocked down through both shRNA and artificial miRNA targeting HNF4A as previously described [20]. They were maintained with 0.5 μg/ml of puromycin and regularly passaged for further analysis.

2.2. Reverse transcription and real-time quantitative PCR

Equal number of cells (8 × 10^5 for HepG2 cells and 2.5 × 10^5 for PLCR/PRF/5 cells) was plated in 12-well plates 24 h before harvesting. Total RNA was extracted from cells with TRizol (Life Technologies) according to the manufacturer's instructions. A total of 1 μg total RNA was reverse transcribed into cDNA with Maxima First Strand cDNA synthesis kit (Life Technologies). The qPCR reactions were performed with JumpStart Taq ReadyMix (Sigma) coupled with EvaGreen dye (Biotium) on the AriaMx Real-time PCR System (Agilent Technologies). The primers used are listed in Supplementary Table 1. The expression was normalized to control transcripts RSP18 and ACTB. The measurement was carried out by two-tailed Student's T-tests.

2.3. Western blot assays

Protein was extracted either simultaneously with total RNA from samples lysed with TRizol (Life Technologies) following the manufacturer's protocol or with RIPA buffer (R0278, Sigma-Aldrich) as previously described [21]. The concentration of extracted protein for each sample was denatured in 1 × NuPAGE LDS Sample Buffer, separated by NuPAGE Novex 4–12% Bis-Tris Protein gel (Life Technologies) and blot transferred onto PVDF transfer membranes by iBlot Dry Blotting System (Life Technologies). Membranes were blocked with 5% BSA in PBS, incubated with the appropriate antibodies targeting HNF4A (SC-374229, Santa Cruz Biotechnology), SCD1 (PA5–19862, Thermo Fisher Scientific) and β-actin (SC-81178,Santa Cruz Biotechnology). HRP-conjugated secondary antibody (SC-2004 or SC-2005, Santa Cruz Biotechnology) was employed to visualize the target protein under a CCD camera from Chem-Doc XRS System (Bio-Rad).

2.4. Chromatin immunoprecipitation

Chromatin immunoprecipitation (ChIP) experiments were carried out as previously described [20]. Briefly, 4 μg of antibody against H3K27ac (ab4729, Abcam) was co- incubated with sonicated chromatin from 1 × 10^6 cells to pull down target regions. For ChIP experiments against HNF4A, chromatin from 3 × 10^6 cells were incubated together with 8 μg HNF4A antibody (SC-374229, Santa Cruz Biotechnology) to precipitate target regions. Each experiment were replicated three times. A standard curve was generated for each target with serial-diluted input DNA as templates. The amounts of DNA pulled down for each target was first normalized to the input based on the standard curve. Then the relative enrichment of target DNA was displayed as fold changes over the negative control region GDCHR12, located in one gene desert region on chromosome 12. The promoter region of HNF1A (HNF1a-pro) was used as the positive control for both H3K27ac and HNF4A. The ChIP followed by quantitative real time PCR (ChIP-qPCR) reactions were performed with JumpStart Taq ReadyMix (Sigma) coupled with EvaGreen dye (Biotium) on the AriaMx Real-time PCR System (Agilent Technologies). All the primer sequences used for qPCR analysis are listed in Supplementary Table 2.

2.5. Allele-specific quantitative PCR

Primers for Allele-specific quantitative PCR (AS-qPCR) were designed based on mismatch amplification mutation assay to accurately quantify single nucleotide mutations [22,23]. As rs55710213 and rs56334587 are adjacent to each other, allele specific forward primers GG_RT: CTGCTTTCCCTCCTCATCACCTC and AA_RT: CTGCTTTCCCTCCTTATCAGCt overlapping with the universal reverse primer AS_RT: AGAAGAGATGAGCAGTGCAGT were designed to validate the allelic imbalance in chromatin binding of HNF4A in ChIP experiments. The ChiP-qPCR reactions were performed with JumpStart Taq ReadyMix (Sigma) coupled with EvaGreen dye (Biotium) on the AriaMx Real-time PCR System (Agilent Technologies). The specificity of the AS-qPCR primers was validated with genomic DNA extracted from samples from the Excellence of Diabetes Research in Sweden (EXODIAB) biobank and different cell lines.

2.6. Luciferase plasmids construction and reporter assay

The promoter and enhancer regions in the SCD1 locus indicative of HNF4A binding were amplified from genomic DNA extracted from HepG2 cells using Phusion High-Fidelity DNA Polymerase (F5308, Thermo Fisher Scientific). The promoter region of SCD1 was inserted into pGL4.10 (Promega) digested with KpnI and EcoRV through the In-focus cloning system (Takara). All the enhancer regions were inserted upstream of the minimal promoter (MP) sequence of pGL4.23 (Promega), which was also digested with KpnI and EcoRV, through the In-focus cloning system (Takara). A three fragments In-focus cloning system with mutations introduced through primer sequences was employed to introduce desired mutations into luciferase constructs. The detailed primer information and cloning method for each luciferase construct is listed in Supplementary Table 3. All the resulting plasmids were verified by Sanger sequencing.
Luciferase plasmids were purified with GenElute Plasmid Miniprep Kit (Sigma Aldrich) and quantified by NanoDrop 2000 (Thermo Fisher Scientific). HepG2 cells were plated one day before transfection in 96-well plates. The confluency was ~70% on transfection. For normal luciferase assay, each well was transfected with 0.3 μl x-tremeGENE HP DNA transfection reagent (Roche) and 100 ng of experimentally firefly luciferase reporter plasmid, and 1 ng of pGL4.74 renilla luciferase reporter vector (Promega) as internal control for monitoring transfection and lysis efficiency. For luciferase assay overexpressing HNF4A, each well was transfected with 0.3 μl X-tremeGENE HP DNA transfection reagent, 40 ng of firefly luciferase reporter plasmid, 50 ng pcDNA3.1-HNF4A and 10 ng of pRL-PM as previously described [24].

Cells were harvested 24 h after transfection and assayed with the Dual-Luciferase Reporter Assay System (Promega) on an Infinite M200 pro reader (Tecan). All the results are expressed directly as the ratio of firefly luciferase activity from experiment plasmids to renilla luciferase activity from control plasmids. All the luciferase experiments came from two independent transfections i.e., independent plasmid preparations and transfections each with three technical replicates. Luciferase values are expressed as averages with error bars representing standard deviations (s.d) from all technical replicates and statistical analyses were performed by two-tailed Student’s t-tests.

3. Results

3.1. The expression of SCD1 was significantly downregulated by HNF4A knockdown

HNF4A preferentially expressed in a number of metabolic organs, including the liver, kidney and intestine [15,30]. We first explored the expression correlation between SCD1 and HNF4A in various cell lines derived from these organs in CellMinerCDB [25]. We found that the expression of SCD1 is significantly correlated with the expression of HNF4A in 134 cell lines in the CCLE database (r = 0.49, p = 2.2e-9), in 100 cell lines in the CTRP database (r = 0.46, p = 1.8e-6), and in 105 cell lines in the GDSC database (r = 0.26, p = 7.8e-3) (Fig. 1A). To confirm the observed expression correlation between SCD1 and HNF4A, we examined the expression changes of SCD1 upon HNF4A knockdown in both HepG2 and PLC/PRF/5 cells [20]. Knockdown of HNF4A significantly decreased the expression of SCD1 in both cells after normalization to reference genes ACTB and RSP18, respectively in quantitative reverse transcription polymerase chain reaction (RT-qPCR) (Fig. 1B, 1C & Supplementary Fig. 1). The decreased expression of both HNF4A and SCD1 induced by HNF4A knockdown was further validated by Western blot in both cells (Fig. 1D, 1E).

3.2. The regulatory regions in the SCD1 locus are enriched with HNF4A binding

The downregulation of SCD1 by HNF4A knockdown could be directly caused by decreased HNF4A binding to the regulatory regions in the SCD1 locus or indirectly by other TFs, e.g. PPARA and PPARG, downregulated by HNF4A knockdown, that further regulate SCD1 expression [31,32]. We evaluated the expression changes of the key TFs and their coactivators involved in lipogenesis induced by HNF4A knockdown [12,33,34]. The expression of SREBF1 was significantly downregulated by HNF4A knockdown in HepG2 cells but not in PLC/PRF/5 cells after normalization with the expression of ACTB and RSP18. Meanwhile, the expression of LXRA and SREBF2 was significantly downregulated by HNF4A knockdown in PLC/PRF/5 cells but not replicated in HepG2 cells (Fig. 2A, 2B & Supplementary Fig. 2). SREBP1, encoded by SREBF1 and LXRA are two of the master transcriptional regulators in hepatic lipogenesis, both of which have also been reported to directly regulate the expression of SCD1 [12,33]. The observed downregulation of SCD1 upon HNF4A knockdown could be partially attributed to the downregulation of SREBF1 in HepG2 cells and to the downregulation of LXRA in PLC/PRF/5 cells.

To assess whether SCD1 is directly regulated by HNF4A, we first identified the active regulatory regions in the SCD1 locus. By exploring H3K27ac ChIP-seq and DNase-seq signals in the SCD1 locus, the promoter of SCD1 (P) and two adjacent enhancers (E1 and E2) were predicted to be highly active in HepG2 cells (Fig. 3A). HNF4A ChIP-seq signals from both liver tissues and HepG2 cells are enriched in the predicted regulatory regions defined by H3K27ac ChIP-seq and DNase-seq. A third HNF4A binding peak, that is present only in the liver tissues and is located directly downstream of the SCD1 gene, is void of H3K27ac ChIP-seq and DNase-seq signals and was named enhancer E3 (Fig. 3A).

We next performed ChIP-qPCR with antibody against H3K27ac in both HepG2 and PLC/PRF/5 cells to validate the identified active regulatory regions in the SCD1 locus. The enrichment of H3K27ac signal at each target region was normalized with the negative control region GDCHR12 as previously described (Fig. 3B) [20,24]. In parallel with the positive control region HNF1α-pro, the predicted enhancers (E1-E3) in the SCD1 locus were highly enriched with H3K27ac signals in both HepG2 and PLC/PRF/5 cells (Fig. 3C). The enrichment of H3K27ac over enhancer E1 and E2 was each validated with two different sets of primers (P1 and P2), which showed similar patterns of fold enrichments over GDCHR12. The H3K27ac signal was significantly enriched in the promoter region of SCD1 in PLC/PRF/5 cells but not in HepG2 cells (Fig. 3C). We then employed the same set of primers to validate the enrichment of HNF4A over each target region in both HepG2 and PLC/PRF/5 cells. Compared with GDCHR12, the positive control region HNF1α-pro was significantly enriched with HNF4A binding in both cells (Fig. 3D). In accordance with the previous observation that HNF4A directly binds to the promoter region of SCD1 in hepatocytes, the promoter region of SCD1 (P) was significantly enriched with HNF4A binding in both HepG2 and PLC/PRF/5 cells (Fig. 3E) [19]. Additionally, all the enhancers (E1-E3) of SCD1 were also significantly enriched with HNF4A binding in both cell lines. (Fig. 3E).
Fig. 1. The expression of *SCD1* was significantly downregulated by HNF4A knockdown. (A) The expression of *SCD1* and HNF4A was highly correlated in cell lines originated from liver, intestine and kidney, in which HNF4A was highly expressed. The gene expression data from CCLE, CTRP and GDSC databases in CellMinerCDB are shown [25]. (B–C) The expression of *SCD1* (C) was significantly downregulated (B) determined by RT-qPCR in both HepG2 cells and PLC/PRF/5 cells. The expression of *SCD1* and HNF4A was normalized to the expression of *ACTB*. *n* = 12 for HepG2 cells and *n* = 16 for PLC/PRF/5 cells. ** *p* < 0.01 calculated by two-tailed Student's *t*-tests. (D–E) The downregulation of HNF4A (D) and SCD1 (E) induced by HNF4A knockdown was further validated by Western blot in both HepG2 cells and PLC/PRF/5 cells. The expression of β-actin was used as the loading control in both cells.
3.3. Identification of key regulatory regions in the SCD1 locus responsive to HNF4A overexpression

As a member of the nuclear receptors, HNF4A regulates gene expression by forming homodimers and binding specific DNA sequences, e.g. the classical direct repeat 1 (DR1) motif, where each “AGGTCA” half-site is separated by a single nucleotide, usually “A” [35,36]. The enrichment of HNF4A binding to regulatory regions can be either caused by HNF4A direct binding or indirectly through protein partners in complex with HNF4A [37]. The identified promoter and three enhancers in the SCD1 locus were all significantly enriched with HNF4A binding. To identify the key regulatory regions in the SCD1 locus mediating the regulatory effect of HNF4A, we combined luciferase reporter assay and in-silico prediction of HNF4A binding using the TRAP tool [38].

The promoter region of SCD1 showed strong promoter activity in luciferase assay (Fig. 4A). However, this region failed to be induced by HNF4A overexpression in luciferase assay even in the presence of the DR1 motif sequences previously identified to be bound by PPARA (Fig. 4A) [10,31]. This suggests that the enrichment of HNF4A signal over the promoter region of SCD1 might be indirectly caused by promoter-enhancer interactions instead of HNF4A direct binding. The enhancer E1 and E2 showed significant enhancer activity, whereas enhancer E3 failed to show any enhancer activity in luciferase assay (Fig. 4B). Upon HNF4A overexpression, the activities of all the three enhancers were significantly induced (Fig. 4B). We then employed the TRAP tool to predict the enrichment of HNF4A binding motifs in all the three enhancers [38]. Multiple HNF4A binding motifs including the DR1 motif were predicted in enhancer E1 and E2 (*P < 0.05), but not in E3 (Fig. 4C). We then focused our analysis on enhancer E1 and E2 to identify how their activities were regulated by HNF4A. We constructed a series of truncation and deletion luciferase constructs for both enhancers to identify the minimal regions responding to HNF4A overexpression (Fig. 4D-4E). For enhancer E1, luciferase construct E1 (1–902) didn’t show enhancer activity, whereas luciferase construct E1 (882–1607) showed significant enhancer activity (Fig. 4D). This indicates that the enhancer activity of E1 was mainly derived from the region from 882 to 1607, where the HNF4A peak summit was located. In accordance with the motif prediction, the enhancer activities of both E1 (1–902) and E1 (882–1607) were significantly induced by HNF4A overexpression. This suggests that the whole E1 region may be needed for the maximal induction of HNF4A. For enhancer E2, the minimal region responsible for the enhancer activity was mapped to the region from 1 to 807, where the HNF4A peak summit was located. In accordance with the motif prediction, the enhancer activities of both E1 (1–902) and E1 (882–1607) were significantly induced by HNF4A overexpression. This suggests that the whole E1 region may be needed for the maximal induction of HNF4A. For enhancer E2, the minimal region responsible for the enhancer activity was mapped to the region from 1 to 807, where the HNF4A peak summit was located. In accordance with the motif prediction, the enhancer activities of both E1 (1–902) and E1 (882–1607) were significantly induced by HNF4A overexpression.
overexpression, indicating this region is also needed for enhancer E2 responding to HNF4A overexpression (Fig. 4E).

3.4. SNPs rs55710213 and rs56334587 reside in a canonical HNF4A motif and regulate the enhancer activity of E2

Genetic variants in the SCD1 locus have been implicated in regulating the expression of SCD1, deciphered by the observed associations between genetic variants in this locus and the SCD1 activity indices.
Fig. 4. The enhancer regions in the SCD1 locus are enriched with canonical HNF4A binding motifs and were highly induced by HNF4A overexpression in luciferase assay. (A) The promoter region of SCD1 (P) showed strong promoter activity, but failed to be induced by HNF4A overexpression in luciferase assay. (B) The enhancer regions E1 and E2 showed significant enhancer activities and were also significantly induced by HNF4A overexpression in luciferase assay. (C) In silico analysis of HNF4A motifs enriched in the enhancer regions by the TRAP tool [38]. (D-E) Fine mapping of HNF4A responsive regions in enhancer E1 (D) and E2 (E) through truncation and deletion luciferase constructs. (F) List of common SNPs (MAF ≥ 1%) in the regulatory regions of the SCD1 locus from dbSNP release 153 [41]. The SNPs directly located in a HNF4A motif are underlined and highlighted in bold. For A-B and D-E, error bars, s.d; n = 6 technical replicates from two independent plasmid extractions and transfections with each transfection had three technical replicates. ns, not significant and ** P < 0.01 calculated by two-tailed Student's t-tests.
3.5. HNF4A preferentially binds to enhancer E2 bearing the AA haplotype (Fig. 5G). A total of 25 SNPs were identified in the regulatory regions of the SCD1 locus (Fig. 4F). Among them, rs55710213 and rs56334587 are separated by one nucleotide in enhancer E2 and predicted to be located in a canonical HNF4A binding motif, which were further selected for functional validation.

SNPs rs55710213 and rs56334587 are in complete linkage in all populations from 1000 Genomes Phase 3 [42]. HNF4A is predicted to preferentially bind to the AA haplotype in comparison with the GG haplotype (Fig. 5A). In accordance with the motif prediction, the enhancer activity of E2 was significantly increased by mutating the G allele of either SNPs to their corresponding A allele in luciferase assay, suggesting that both SNPs are functional in regulating the enhancer activity of E2 (Fig. 5B). The significant difference in enhancer activity between haplotypes was further replicated in the truncation luciferase construct E2 (786–1216) (Fig. 5C). We next evaluated whether E2 bearing different haplotypes responds differently to HNF4A overexpression in luciferase assay. The enhancer activity of E2 with either of the haplotypes was significantly induced by HNF4A overexpression, probably due to the presence of multiple HNF4A binding sites in E2. However, the significant difference in enhancer activity between haplotypes were not diminished by HNF4A overexpression (Fig. 5D). In addition to the HNF4A binding site containing rs55710213 and rs56334587, another canonical HNF4A bindings site was predicted in the region from 841 to 855 in the truncation luciferase construct E2 (786–1216) (Fig. 5E). To better delineate the haplotypic differences in response to HNF4A overexpression, we further mutated the second HNF4A binding site in luciferase construct E2 (786–1216) and observed significantly decreased enhancer activity of the luciferase construct bearing either of the haplotypes (Fig. 5F). The luciferase construct with both HNF4A binding sites mutated (GG-HNFmut) had the lowest enhancer activity, whereas the luciferase construct with both of the HNF4A binding sites being conserved (AA) had the highest enhancer activity. The luciferase constructs with either of the HNF4A binding sites mutated (GG and AA-HNFmut) had comparable intermediate enhancer activities (Fig. 5F). Accordingly, mutation of both HNF4A binding sites significantly decreased its response to HNF4A overexpression compared with either of the HNF4A binding sites being conserved or both of the HNF4A binding sites being conserved, indicating both of the predicted HNF4A binding sites are required for maximal HNF4A induction (Fig. 5G).

3.6. Linkage disequilibrium between rs55710213, rs56334587 and the reported tag SNPs in the SCD1 locus

As surrogate measures of SCD1 activity, the SCD1 activity indices have been positively associated with the prevalence of hypertension, metabolic syndrome, FCHL and several other obesity related traits [7–9,39,43]. Both of our in vitro and in vivo assays suggest that rs55710213 and rs56334587 in a haplotype are functional in regulating the enhancer activity of E2, which may affect the SCD1 activity indices and further disease prevalence by directly regulating the expression of SCD1. This hypothesis is supported by the observation that genetic variants in the SCD1 locus have been associated with not only the SCD1 activity indices but also several other phenotypic traits including obesity, insulin sensitivity, metabolic syndrome and inflammation [39,40,43,44].

To verify if rs55710213 and rs56334587 are the causative variants underlying these associations, we further investigated whether the two SNPs are in linkage disequilibrium (LD) with the reported tag SNPs associated with various phenotypes. The minor allele of rs7849 in linkage with the GG haplotype of rs55710213 and rs56334587 (D′ = 1 and r2 = 0.12 in SweGen population) was associated with higher insulin sensitivity and less waist circumference in an elderly Swedish population (Fig. 7A) [40,45]. The major allele of rs1502593 in LD with the GG haplotype of rs55710213 and rs56334587 (D′ = 0.81 and r2 = 0.54 in EUR population) has been associated with lower prevalence ratios of metabolic syndrome in Costa Rican adults and lower odds ratio of obesity in a Spanish population from the Pizarra study (Fig. 7B) [39,43]. These observations supported the hypothesis that carriers of the GG haplotype have a lower SCD1 activity indices due to reduced expression of SCD1, which decreased the synthesis of MUFAs and further affected disease progression.

The associations between genetic variants in the SCD1 locus and the SCD1 activity indices have been reported in two studies so far, with borderline and inconclusive results. The minor alleles of rs508384, rs10083463, rs2167444 and rs7849 were associated with both a higher SCD1 activity index measured by the 16:1/16:0 ratio and a lower activity index measured by the 18:1/18:0 ratio in the serum in the Spanish population from the Pizarra study [39]. Due to the low allele frequencies of the minor alleles of all the four reported SNPs, the correlation (r2) between haplotypes of rs55710213 and rs56334587 and alleles of the four SNPs are very low that ranged from 0.03 to 0.11 in the EUR population (Fig. 7B). However, the minor alleles of all the four reported SNPs are predominantly co-localized with the GG haplotype, indicated by the high D′ value that ranged from 0.96 to 1 in the EUR population. This may suggest that the less active GG haplotype of rs55710213 and rs56334587 may contributed to the observed higher 16:1/16:0 ratio and lower 18:1/18:0 ratio in carriers bearing the minor alleles of the four SNPs. This hypothesis is partially supported by the other study that performed in the serum in an elderly Swedish population. The major
Fig. 5. The HNF4A-motif disrupting SNPs rs55710213 and rs56334587 are functional in regulating the enhancer activity of E2. (A) The SNPs rs55710213 and rs56334587 are adjacent to each other and located in a canonical HNF4A binding site. The location of this HNF4A binding site is marked relative to the original luciferase construct of E2. (B) Enhancer activity of E2 bearing different alleles of rs55710213 and rs56334587 in luciferase assay. (C) Luciferase constructs bearing the AA haplotype showed significant higher enhancer activity compared with the constructs bearing the GG haplotype. (D) Different responses of luciferase constructs bearing either the AA or GG haplotype to HNF4A overexpression in luciferase assay. (E) The location of the second HNF4A binding site predicted in enhancer E2. The mutations introduced into this HNF4A binding site in luciferase constructs are highlighted in red. (F-G) Introducing mutations into the second HNF4A binding site significantly decreased the enhancer activity of E2 and its response to HNF4A overexpression in luciferase assay. For B–D and F–G, error bars, s.d; n = 6 technical replicates from two independent plasmid extractions and transfections with each transfection had three technical replicates. ** P < 0.01 calculated by two-tailed Student’s t-tests.
allele of rs3071 in high LD with the GG haplotype of rs55710213 and rs56334587 inДЕчIP-seq experiments with antibody against HNF4A in liver tissue (GEO: GSE96260). (B) Validation of AS-qPCR assay targeting rs55710213 and rs56334587 with genomic DNA from samples that bearing different haplotypes at this locus as template. (C-D) HNF4A favors binding to the AA haplotype relative to the GG haplotype as determined by ChIP followed by AS-qPCR (C) and ChIP followed by PCR amplification and Sanger sequencing (D) in HepG2 cells. The input DNA was used as the control. Error bars, s.d; n = 6 technical replicates. ** P < 0.01 calculated by two-tailed Student’s t-tests.

Fig. 7. Linkage disequilibrium between rs55710213, rs56334587 and the previously reported tag SNPs in the SCD1 locus. The standard (D'/LOD) color scheme was applied for the plots. The number in each well denotes $r^2$ value between pair of SNPs in the SweGen (A) and European (EUR in 1000 Genomes Phase 3) population (B).

4. Discussion

We here demonstrated that HNF4A regulates SCD1 expression by directly binding to the key regulatory regions in the SCD1 locus. The enhancers both located upstream of the SCD1 gene (E1) and in intron 5 of SCD1 (E2) are the key cis-regulatory elements for HNF4A and in intron 5 of SCD1 (E2) are the key cis-regulatory elements for HNF4A binding. rs55710213 and rs56334587 located in a consensus HNF4A motif in enhancer E2 are functional. The AA haplotype of rs55710213 and...
rs56334587 is preferentially bound by HNF4A, which further increased the enhancer activity of E2. We proposed that the increased enhancer activity of E2 may increase the expression of SCD1 in response to HNF4A, which further increase the efficiency in the synthesis of MUFAs (Fig. 8).

Even though HNF4A is highly expressed in both liver and gastrointestinal organs, its perturbations in various diseases have been most extensively investigated in the context of liver diseases [16]. Dysregulated HNF4A have been associated with many liver diseases, including alcoholic liver disease, NAFLD, fibrosis and cirrhosis, viral hepatitis and hepatocellular carcinoma [16]. These associations may partially be attributed to the fact that HNF4A is a central regulator of both glucose and lipid metabolism in the liver [15,16]. As SCD1 encodes one of the key enzymes in lipogenesis, the identification of SCD1 as a direct target of HNF4A in this study suggested that dysregulated HNF4A may significantly alter the lipogenesis pathway in the liver, which may further contribute to the prevalence of many diseases. There are several lines of evidence that support this hypothesis. First, the ligand binding domain (LBD) of HNF4A is known to be constitutively bound by fatty acids with binding of SCD1 to the LBD activated the activity of HNF4A and binding of unsaturated fatty acids repressed its activity [46,47]. SCD1 catalyze the synthesis of fatty acids from SFAs to MUFAs and both the precursor and product of SCD1 are the natural ligands of HNF4A [47]. HNF4A may act as a metabolic sensor that can finely tune the expression of SCD1 in response to a variety of nutritional and hormonal signals demanding different amounts of fatty acids. Second, several of the crucial TFs regulating hepatic lipogenesis, including SREBP1, CHREBP1, and PPARA/PPARG are directly regulated by HNF4A [48,49]. SCD1 is known to be regulated by LXR, SREBP1 and PPARA by direct binding to its promoter region [10,12]. The observed downregulation of SCD1 upon HNF4A knockdown may be partially caused by the downregulation of SREBP1 in HepG2 cells and by the downregulation of LXR in PLC/PRF/5 cells in our study (Fig. 2). Third, direct evidence from murine models proved that HNF4A exerted important function in lipogenesis. Knockdown of HNF4A or disrupting the binding of HNF4A in mouse hepatocytes significantly inhibited the hepatic de novo lipogenesis pathway [32,50]. Finally, MODY1 patients bearing inactivating HNF4A mutations have reduced levels of serum TGs and cholesterol and display defects in the expression of genes in lipid metabolism [51]. Interestingly, mice model with either hepatocyte-specific HNF4A knock out or HNF4A knockdown also displayed reduced plasma levels of TGs and cholesterol, which is caused by both reduced very low-density lipoprotein secretion and hepatic lipogenesis [32,52].

The identification of rs55710213 and rs56334587 as functional regulatory variants in the SCD1 locus suggested that these two SNPs may affect the SCD1 enzyme activity and subsequently contribute to disease development. However, it should be noted that the association between genetic variants and the SCD1 activity indices have been reported in only two studies [39,40]. The observed weak associations in these two studies can be partially attributed to the relatively small cohort with 824 participants in the Pizarra study and 489 participants in the Swedish study [39,40]. Additionally, both the SFA precursors and MUFA products of SCD1 are the predominant fatty acids in human diet [53]. The intake of dietary fat have been proved to play a significant impact on the SCD1 activity indice [54]. Finally, the SCD1 activity indice were evaluated in specific components of serum or plasma lipids [4,39,40]. For example, the SCD1 activity indice were evaluated from serum phospholipids in the Pizarra study, whereas the SCD1 activity indice were evaluated from serum cholesterol esters in the Swedish study [39,40]. Measuring SCD1 activity indices in the serum may fail to accurately represent liver-specific changes in SCD1 activity. Further replicative studies in larger cohorts will be helpful in clarifying these mentioned uncertainties.

The SCD1 gene is located on 10q24.31 in humans. Genetic variants in this region have been associated with various phenotypic traits that closely related to fat metabolism, e.g. body mass index, type 2 diabetes, total cholesterol, HDL cholesterol and many fatty-acids derived metabolites in multiple GWAS [55]. Among the associated fatty-acid derived metabolites, variants in this region were most frequently associated with changing levels of palmitoleic-acid (C16:1) derived metabolites (Supplementary Table 4). Palmitoleic acid (C16:1) is one of the major MUFA catalyzed by SCD1 with palmitic acid as substrate. This makes SCD1 the most promising candidate gene underlying these associations in this region. However, due to the high rate of recombination surrounding the SCD1 locus in human population, the tag SNPs (rs603424 in PKD2L1 and rs10883511 in HIF1AN) reported in GWAS are in linkage equilibrium with our identified functional SNPs rs55710213 and rs56334587 [56]. This implies that there are other functional regulatory variants, in addition to rs55710213 and rs56334587, involved in SCD1 regulation in this region. Moreover, the regulatory variants in linkage with the tag SNPs might be involved in the regulation of other candidate genes in this region, e.g. HIF1AN, which is a regulator of lipid metabolism [57].

In this study, due to the absence of exonic variants in the SCD1 gene in HepG2 cells, the allelic imbalance in SCD1 expression associated with
the haplotypic imbalance of HNF4A binding to rs55710213 and rs56334587 in enhancer E2 could not be determined. We have also explored several other liver-oriented cell lines including PLC/PRF/5 and Hep3B and found both of the cell lines homozygous at rs55710213 and rs56334587 with either GG haplotype in PLC/PRF/5 cells or AA Hep3B and found both of the cell lines homozygous at rs55710213 and
be designed in the future. The identified functional regulatory SNPs will provide new clues to the altered metabolic pathways in many re
explored several other liver-oriented cell lines including PLC/PRF/5 and
Supervision, Writing-review and editing.
Claes Wadelius

CRediT authorship contribution statement

Gang Pan: Conceptualization, Formal analysis, Investigation, Methodology, Resources, Writing-original draft, Writing-review and editing.
Marco Cavalli: Writing-review and editing.
Claes Wadelius: Conceptualization, Funding acquisition, Methodology, Supervision, Writing-review and editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This study was supported grants from Swedish Cancer Foundation (CAN 2018/849), the Swedish Diabetes Foundation (DIA 2017-269), the Family Ermfors Foundation and EXODIAB to CW.

References
