Genetic determinants of cholesterol and energy metabolism : implications for cardiometabolic health
Blauw, L.L.

Cover Page

Universiteit Leiden

5

Exome-wide analysis for rare variants identifies *ABCA6* as determinant of circulating CETP concentration

Lisanne L. Blauw, Ruifang Li-Gao, Raymond Noordam, Jan B. van Klinken, Yanan Wang, Zhuang Li, Renée de Mutsert, Stella Trompet, Jimmy F.P. Berbée, Diana van Heemst, Dennis O. Mook-Kanamori, Frits R. Rosendaal, J. Wouter Jukema, Patrick C.N. Rensen, Ko Willems van Dijk

5

Abstract

5

Introduction

Recently, we performed a genome-wide association analysis (GWAS) to identify common genetic determinants of circulating cholesteryl ester transfer protein (CETP), an approach by which rare variants are missed. Identification of novel coding variants may provide insight in the pathways that regulate CETP metabolism, and may contribute to our understanding of the biological function of CETP in lipid metabolism and cardiovascular disease. Therefore, in the present study, we aimed to identify rare coding variants in exonic regions that determine CETP concentration.

Methods

We performed an exome-wide gene-based aggregation analysis in 6,094 participants from the Netherlands Epidemiology of Obesity (NEO) study to identify rare coding variants (minor allele frequency <0.03) in exonic regions that are associated with serum CETP concentration.

Results

We identified the *ABCA6* gene to be exome-wide significantly associated with CETP concentration (P=9.4×10^{-7}). The driving variant for the *ABCA6* genetic region was rs77542162 (MAF=0.028) with a per-allele effect of 0.19 μg/mL on circulating CETP concentration, while the mean (SD) CETP concentration in the study population was 2.48 (0.66) μg/mL.

Conclusion

The *ABCA6* gene was identified as a novel determinant of circulating CETP, which indicates that abundance or functionality of the ABCA6 protein regulates CETP concentration. Future research is warranted to elucidate the underlying biological pathways that can explain this association.

Introduction

Cholesteryl ester transfer protein (CETP) is able to transfer cholesteryl esters from high-density lipoproteins (HDL) towards apolipoprotein B (ApoB) containing triglyceride-rich lipoproteins, mainly very low-density lipoproteins (VLDL). In exchange, triglycerides are transferred from VLDL towards triglyceride-poor particles, which are both HDL and low-density lipoproteins (LDL).[1] As such, CETP contributes to an atherogenic lipoprotein profile (i.e. high LDL-cholesterol (C)/HDL-C ratio), as has been extensively studied in both humans and in mice transgenic for human CETP.[2,3] Therefore, inhibition of CETP as therapeutic strategy to attenuate dyslipidaemia and ultimately prevent the development of cardiovascular disease (CVD) has been extensively tested.

To elucidate the genetic basis of CETP, we recently performed a genome-wide association study (GWAS) on serum CETP concentration in the Netherlands Epidemiology of Obesity (NEO) study, and showed that CETP concentration is highly genetically determined by three independent loci with respectively driving single nucleotide polymorphisms (SNPs) rs247616, rs12720922 and rs1968905, which are all mapped to the *CETP* gene [Blauw LL, et al. Submitted]. This GWAS was specifically designed to identify common genetic variants in relation to CETP concentration, an approach by which rare variants that may have larger effects on serum CETP concentration than common variants are not taken into account.

Identification of novel coding gene regions that determine CETP concentration may provide insight in the pathways that regulate CETP production and clearance. Elucidation of these pathways may contribute to our understanding of the biological function of CETP in lipid metabolism and CVD. This is especially relevant in light of the recent disclosure of the results from the REVEAL trial, which showed that the CETP inhibitor anacetrapib met its primary endpoint, i.e. a reduction in major coronary events.[4] Therefore, in the present study, we aimed to identify rare coding variants (i.e. minor allele frequency (MAF) <0.03) in exonic regions that are jointly associated with serum CETP concentration by performing an exome-wide gene-based aggregation analysis.

Materials and Methods

Study design and population

The NEO study is a population-based prospective cohort study of men and women aged between 45 and 65 years. From the greater area of Leiden, The Netherlands, all inhabitants with a self-reported body mass index (BMI) of 27 kg/m^2 or higher were eligible to participate. In addition, inhabitants from one nearby municipality (Leiderdorp, The Nether-

lands) in the same age group were invited to participate regardless of their BMI, forming a reference population for BMI distribution. In total, 6,671 participants were included from September 2008 until September 2012. Participants visited the NEO study center for extensive physical examination. Venous blood samples were obtained from the antecubital vein after a 10 hour overnight fast. Research nurses recorded current medication use by means of a medication inventory. Prior to the study visit, participants completed questionnaires at home with respect to demographic, lifestyle, and clinical information.

The NEO study was approved by the medical ethics committee of the Leiden University Medical Center (LUMC), and all participants gave their written informed consent. Detailed information about the study design and data collection has been described elsewhere.[5]
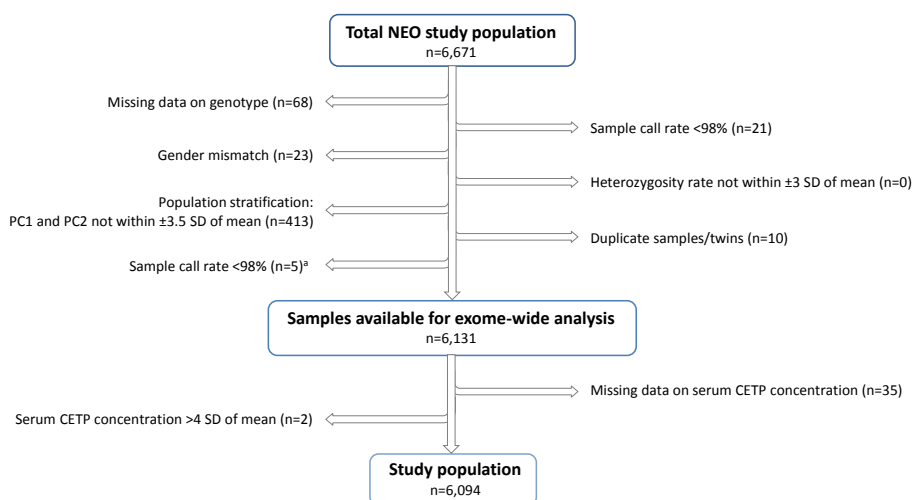
Figure 5.1: Quality control steps for the genome-wide analysis and exclusion criteria for the discovery and replication cohort from the Netherlands Epidemiology of Obesity (NEO) study.

[a] Sample call rate was checked a second time, as it can change after removing samples on the basis of prior quality control steps.
CETP, cholesteryl ester transfer protein; PC, principal component.

Biochemical analyses

After centrifugation, aliquots of plasma and serum were stored at -80°C. From 11 April until 16 July 2014 CETP concentrations were measured in serum that had undergone one previous freeze-thaw cycle with enzyme-linked immune sorbent assay (ELISA) kits according to the manufacturer's instructions (DAIICHI CETP ELISA, Alpco, Salem, USA; coefficient of variation 11.7%). Participants with missing data on serum CETP concentration (n=35) and participants with a serum CETP concentration beyond four SD from the mean (n=2)

were excluded from analyses.

Fasting serum total cholesterol and triglyceride concentrations were measured with enzy-matic calorimetric assays (Roche Modular P800 Analyzer, Roche Diagnostics, Mannheim, Germany; CV <2% and CV <3%, respectively) and fasting serum HDL-C concentrations with third generation homogenous HDL-C methods (Roche Modular P800 Analyzer, Roche Diagnostics, Mannheim, Germany; CV <3%). Fasting LDL-C concentrations were calcu-lated using the Friedewald equation.[6]

Genotyping

DNA was isolated from venous blood samples. Genotyping was performed using the Illu-mina HumanCoreExome-24 BeadChip (Illumina Inc., San Diego, California, United States of America), including 265,919 exome-focussed markers. Participants were excluded in the process of quality control when 1) the sample call rate was <98%, 2) there was a sex mismatch, 3) heterozygosity rate was not within ±3 SD of mean heterozygosity rate, 4) participants widely diverged based on the first two principal components (±3.5 SD), and 5) samples were duplicates. The present study population consists of 6,094 individuals (Figure 5.1).

Genetic variants were excluded when genotype call rate was <98%, and when variants were not in Hardy-Weinberg equilibrium (P-value $<1×10^{-6}$). In the present study, we re-stricted to genetic variants, which were located in exonic regions, which had a MAF<0.03, and which had deleterious effects (i.e. splice donor, splice acceptor, stop gained, stop lost, initiator codon or missense).

Gene-based aggregation analysis

We performed a gene-based aggregation analysis to identify rare coding variants (i.e. MAF <0.03) that are jointly associated with serum CETP concentration. With this approach, a gain in statistical power is achieved by evaluating the cumulative effects of multiple genetic variants in a gene region.[7] Two different gene-based aggregation methods were applied to test for different assumptions about the underlying genetic model. The first method, i.e. the combined multivariate and collapsing (CMC) method,[8] is a Burden test, which assumes that 1) the variants in a genetic region are causal and 2) that effects of these variants on the outcome all have the same direction and magnitude. The second method, i.e. the sequence kernel association test (SKAT),[9] is a variance-component test. This test assumes that the effects of the variants on the outcome can have opposite directions (i.e. both positive and negative), and does not make assumptions about causality. We used RAREMETAL version 4.13.9[10] to perform the analyses with the CMC method and SKAT.

To account for the multiple testing of 12,157 genes, a Bonferroni-corrected threshold of $P<4.1\times10^{-6}$ was regarded exome-wide significant.

To identify the driving variants within the identified gene regions, conditional analysis was used. The variant with the lowest P-value was selected, and subsequently the association analysis was adjusted for this variant. If the joint association thereafter showed no statistical significant result anymore, the variant that was adjusted for was proven to be the driving variant within the gene region. Otherwise, the analysis was additionally adjusted for the next variant with the lowest P-value. This procedure was repeated until no more exome-wide significant variants were identified.

We reported for the CMC test, both the joint effect size for the identified gene region and the single effect size of the driving variant. For SKAT only the single effect size of the driving variant was reported, as a joined effect size cannot be generated form this method.

In the case that a genetic region was identified in near proximity of the *CETP* gene, we tested whether the association between the driving variant of this genetic region and serum CETP concentration was explained by one of the common CETP SNPs that determine CETP concentration (i.e. rs247616, rs12720922 and rs1968905, as previously identified with GWAS [Blauw LL, Li-Gao R, et al. Under review]). This was done by adding the three common *CETP* SNPs as covariates to a linear regression model for the association between the rare driving variant and CETP concentration. If the association between the identified rare variant and CETP concentration lost statistical significance after adjustment for the three common *CETP* variants, we regarded the rare variant a proxy for the common CETP variants, rather than an independent determinant of circulating CETP concentration.

5

Results

The characteristics of the 6,094 study participants are described in Table 5.1. Median age was 56 years and 52% of participants were women. The mean (SD) CETP concentration was 2.48 (0.66) μg/mL. With the gene-based aggregation analysis, we identified regions in the *ABCA6* and *NLRC5* genes to be exome-wide significantly associated with serum CETP concentration (Table 5.2).

Table 5.1: Characteristics of the study population (n=6,094).

Characteristics	
Age (year)	56 (51, 61)
Women (n, %)	3,168 (52.0%)
Body mass index (kg/m^2)	29.5 (27.3, 32.3)
Lipid-lowering drug users	963 (15.8%)
Fasting serum concentrations	
CETP (μg/mL)	2.48 (0.66)
Total cholesterol (mmol/L)	5.66 (1.07)
LDL-C (mmol/L)	3.56 (0.98)
HDL-C (mmol/L)	1.43 (0.41)
Triglycerides (mmol/L)	1.25 (0.88, 1.76)

Results are presented as median (inter quartile range) for not normally distributed data, mean (SD) or number (percentage).
[a] Missing data: n=14 for total cholesterol, LDL-C, HDL-C and triglycerides concentrations.
CETP, cholesteryl ester transfer protein; C, cholesterol; HDL, high-density lipoprotein; LDL, low-density lipoprotein.

5

Twelve variants that mapped to the *ABCA6* gene on chromosome 17 were jointly associated with a 0.16 μg/mL higher CETP concentration (joint effect P=9.4×10^{-7}), using the CMC method. The analyses performed with SKAT confirmed this association (joint effect P=3.2×10^{-7}). Of the 12 different genetic variants, the rs77542162 polymorphism was the driving variant of this region with a per-allele increase in CETP concentration of 0.19 μg/mL. This driving variant had a MAF of 0.028.

Twenty variants that mapped to the *NLRC5* gene were jointly associated with a higher CETP concentration, when tested with SKAT only (joint effect P=2.8×10^{-9}). The driving variant of this genetic region was the rs1672867 polymorphism (MAF=0.017). This variant was associated with a per-allele increase in CETP concentration of 0.26 μg/mL. Analyses performed with the CMC method showed no joint association for the *NLRC5* region (joint effect P=0.79). NLRC5 is located directly downstream of the *CETP* gene on chromosome 16. When the association between the *NLRC5* driving variant rs1672867 and serum CETP concentration was adjusted for three common *CETP* variants that determine CETP concentration, the association between rs1672867 and CETP concentration lost statistical significance (P=0.089).

Table 5.2: Summary statistics of the rare variants that were associated with serum CETP concentration, as identified with exome-wide gene-based aggregation analysis using either the combined multivariate and collapsing (CMC) method or the sequence kernel association test (SKAT) (n=6,094).

Chr	Gene	Number of aggregated variants	Amino acid change (minor allele count)	CMC method[a]				SKAT[b]		
				Joint effect size (μg/mL)	P-value	Conditioned P-value[c]	Effect size driving variant[d](μg/mL)	P-value	Conditioned P-value[c]	Effect size driving variant[d](μg/mL)
17	*ABCA6*	12	p.Thr1581Ile (2), p.Leu1556Ile (1), p.Ile1494Thr (4), p.Thr1441Met (1), Ala1379Thr (62), p.Cys1359Arg (336), p.Gly1222Val (1), p.Cys1216Gly (3), p.Cys1059Tyr (9), p.Met686Ile (6), p.Arg660His (19), p.Asn565Ser (3) p.Arg152Gln (1), p.Ala155Thr (7), p.Ser210Leu (316), p.Ala382Ser (1), p.Thr470Ile (1), p.Tyr540Cys (5), p.Gly835Cys (11), p.His857Arg (78),	0.16	9.4×10^{-7}	0.58	0.19	3.2×10^{-7}	0.84	0.19
16	*NLRC5*	20	p.Asn907Asp (210), p.Ala1033Thr (5), p.Val1049Met (1), p.Val1217Met (10), p.Asp1228Asn (280), p.Lys1297Asn (337), p.Gly1317Arg (15), p.Val1367Ile (5), p.Arg1401His (72), p.Pro1667Gln (23), p.Arg1705His (26), p.Ser1817Gly (1)	0.005	0.79	0.05	0.26	2.8×10^{-9}	0.01	0.26

The driving variant is underlined. A P-value $<4.1 \times 10^{-6}$ is regarded exome-wide significant.
[a] The CMC method assumes that variants are causal and have the same effect direction.
[b] SKAT assumes that variants can have different effect directions. A joint effect size can therefore not be generated from the model, only a P-value.
[c] P-value conditioned on the driving variant.
[d] Per-allele effect.
CETP, cholesteryl ester transfer protein; Chr, chromosome.

5

Discussion

In this study, we identified *ABCA6* as a novel determinant of circulating CETP concentration in a Dutch population-based cohort of 6,094 participants. In an exome-wide gene-based aggregation analysis for rare variants, 12 variants mapped to *ABCA6* were jointly associated with CETP concentration. The per-allele effect of the driving variant for the *ABCA6* region on CETP concentration was relatively large (i.e. 0.19 μg/mL) compared to the mean CETP concentration of 2.48 (0.66) μg/mL, although not as large as the effect size that we recently found for a GWAS-identified common genetic determinant of CETP concentration (i.e. 0.35 μg/mL) [Blauw LL, Li-Gao R, et al. Under review]. Notably, GWAS-identified common variants that determine CETP concentration were all mapped to the *CETP* gene, while we identified no rare *CETP* variants in the present study. This may indicate that within the *CETP* gene mainly common variation determines CETP expression.

In addition to *ABCA6*, the *NLRC5* gene, which is located directly downstream of *CETP* on chromosome 16, was identified as a determinant of circulating CETP concentration. However, this association lost statistical significance after adjustment for the three common *CETP* SNPs that strongly determine circulating CETP concentrations, as we previously identified with a GWAS [Blauw LL, Li-Gao R, et al. Under review]. Since the three common *CETP* SNPs were very strongly associated with serum CETP concentration (lowest P-value 3.98×10^{-100}), it is presumable that even variants in low linkage disequilibrium with these common variants may show significant effects on CETP concentration. The loss of the association between *NLRC5* and CETP concentration after adjustment indicates that *NLRC5* is not an independent determinant of circulating CETP, but rather a proxy variant for the common CETP SNPs.

We identified 12 variants mapped to *ABCA6* that were jointly associated with CETP concentration. Since we used two distinct tests (i.e. CMC method and SKAT) with different underlying assumptions about causality and effect directions of the aggregated variants in the gene region,[7] we were able to compare the results of these two tests to shed light on the underlying genetic model of the identified joint variants. For the *ABCA6* variants, both tests showed comparable results, which suggests that all variants in the identified *ABCA6* region have unidirectional effects on CETP concentration.[11] Of the 12 jointly identified *ABCA6* variants , rs77542162 was shown to be the driving variant of this gene region. Interestingly, in 2015, this same *ABCA6* exonic variant was identified to determine LDL-C and total cholesterol concentrations in the Dutch population independent of other loci,[12] which was thereafter replicated in an independent population.[13] The cysteine to arginine change at position 1359 induced by the rs77542162 missense variant is predicted by PolyPhen-2[14] to be 'probably damaging' with a probabilistic score of 0.999. This implies that this

missense mutation is very likely to damage the structure and reduce the function of the ABCA6 protein. [12,14] Taken together, the results of the present study indicate that abundance or functionality of the ABCA6 protein influence circulating CETP concentrations.

It is interesting to speculate on how ABCA6 may influence the circulating CETP concentration. In 2001, *ABCA6* was cloned and sequenced from human macrophages and identified as a novel member of the ATP binding cassette (ABC) A transporter subfamily, [15] of which the members are involved in cellular transmembrane transport of sterols. [16] Although a relation between ABCA6 and CETP has not been described before, striking analogies exist between the two proteins. First, expressions of both *ABCA6* and *CETP* in humans are highest in liver tissue. [14,15] For CETP, specifically resident liver macrophages (i.e. Kupffer cells) were identified as the main determinants of circulating CETP concentration. [14] Although the main hepatic cell type expressing *ABCA6* is currently unknown, *ABCA6* is also expressed by macrophages. [17] Second, ABCA6[17] and CETP[18] are both upregulated during monocyte-to-macrophage differentiation, which suggests that they may be involved in macrophage polarization and immunity. Third, expression of both genes is regulated in response to sterols, albeit that *ABCA6* is downregulated[17] and *CETP* is upregulated. [19] Although the function of ABCA6 is far from completely understood, ABCA6 has been proposed to be involved in cellular lipid and sterol transport. [15,20,21] As *CETP* expression is regulated by sterol derivatives via a liver X receptor α (LXRα) responsive element in its promoter region, [19,22] it is tempting to speculate that ABCA6 determines CETP production via intracellular regulation of lipid and sterol trafficking.

The MAF of the driving variant of the *ABCA6* region (i.e. rs77542162) was 0.028 in the present Dutch study population. Previously, it has been shown that the minor allele of rs77542162 is enriched in the Dutch population (i.e. MAF=0.030; Genome of the Netherlands (GoNL) population) by 3.65-fold, compared to the world population (i.e. MAF=0.008; 1000 Genomes Project (1KG) population), which implies that there has been genetic drift for this locus in the Netherlands. [12] Due to the enrichment of this variant in the study population, we benefited from an improved power to identify this rare variant as a determinant of CETP concentration. Although we did not replicate the present findings in an independent non-Dutch population, it is interesting to note that the results of a previous study, which used the Dutch population to identify the rs77542162 as a determinant of LDL-C and total cholesterol concentration, were robustly replicated in independent international populations. [12,13] This suggests that results for rs77542162 from a specific homogeneous population may well-translate to a worldwide scale. [12]

In conclusion, in the present study we used exome-wide analysis to show that rare variants mapped to *ABCA6* are associated with circulating CETP concentration. This finding is

5

completely novel, as the *ABCA6* gene has never been described before in relation to CETP. Future research is warranted to elucidate the underlying biological pathways that explain this novel association. Hypothetically, abundance or functionality of the ABCA6 protein may determine CETP production via the regulation of intracellular lipid and sterol trafficking.

5

References

[1] A. R. Tall. Plasma cholesteryl ester transfer protein. *J Lipid Res*, 34(8):1255–74, 1993.

[2] M. J. Chapman, W. Le Goff, and M. Guerin et al. Cholesteryl ester transfer protein: at the heart of the action of lipid-modulating therapy with statins, fibrates, niacin, and cholesteryl ester transfer protein inhibitors. *Eur Heart J*, 31(2):149–64, 2010.

[3] M. Westerterp, C. C. van der Hoogt, and W. de Haan et al. Cholesteryl ester transfer protein decreases high-density lipoprotein and severely aggravates atherosclerosis in APOE*3-Leiden mice. *Arterioscler Thromb Vasc Biol*, 26(11):2552–9, 2006.

[4] The HPS3/TIMI55–REVEAL Collaborative Group. Effects of anacetrapib in patients with atherosclerotic vascular disease. *N Engl J Med*, 377:1217–1227, 2017.

[5] R. de Mutsert, M. den Heijer, and T. J. Rabelink et al. The Netherlands Epidemiology of Obesity (NEO) study: study design and data collection. *Eur J Epidemiol*, 28(6):513–23, 2013.

[6] W. T. Friedewald, R. I. Levy, and D. S. Fredrickson. Estimation of the concentration of low-density lipoprotein cholesterol in plasma, without use of the preparative ultracentrifuge. *Clin Chem*, 18(6):499–502, 1972.

[7] S. Lee, G. R. Abecasis, and M. Boehnke et al. Rare-variant association analysis: study designs and statistical tests. *Am J Hum Genet*, 95(1):5–23, 2014.

[8] B. Li and S. M. Leal. Methods for detecting associations with rare variants for common diseases: application to analysis of sequence data. *Am J Hum Genet*, 83(3):311–21, 2008.

[9] M. C. Wu, S. Lee, and T. Cai et al. Rare-variant association testing for sequencing data with the sequence kernel association test. *Am J Hum Genet*, 89(1):82–93, 2011.

[10] S. Feng, D. Liu, and X. Zhan et al. RAREMETAL: fast and powerful meta-analysis for rare variants. *Bioinformatics*, 30(19):2828–9, 2014.

[11] S. Lee, M. C. Wu, and X. Lin. Optimal tests for rare variant effects in sequencing association studies. *Biostatistics*, 13(4):762–75, 2012.

[12] E. M. van Leeuwen, L. C. Karssen, and J. Deelen et al. Genome of The Netherlands population-specific imputations identify an ABCA6 variant associated with cholesterol levels. *Nat Commun*, 6:6065, 2015.

[13] A. Helgadottir, S. Gretarsdottir, and G. Thorleifsson et al. Variants with large effects on blood lipids and the role of cholesterol and triglycerides in coronary disease. *Nat Genet*, 48(6):634–639, 2016.

[14] I. A. Adzhubei, S. Schmidt, and L. Peshkin et al. A method and server for predicting damaging missense mutations. *Nat Methods*, 7(4):248–9, 2010.

[15] W. E. Kaminski, J. J. Wenzel, and A. Piehler et al. ABCA6, a novel a subclass ABC transporter. *Biochem Biophys Res Commun*, 285(5):1295–301, 2001.

[16] F. Quazi and R. S. Molday. Lipid transport by mammalian ABC proteins. *Essays Biochem*, 50(1):265–90, 2011.

[17] A. Piehler, W. E. Kaminski, and J. J. Wenzel et al. Molecular structure of a novel cholesterol-responsive A subclass ABC transporter, ABCA9. *Biochem Biophys Res Commun*, 295(2):408–16, 2002.

[18] D. Lakomy, C. Rébé, and A. Sberna et al. Liver X Receptor–Mediated Induction of Cholesteryl Ester Transfer Protein Expression Is Selectively Impaired in Inflammatory Macrophages. *Arterioscler Thromb Vasc Biol*, 29(11):1923–1929, 2009.

[19] Y. Luo and A. R. Tall. Sterol upregulation of human CETP expression in vitro and in transgenic mice by an LXR element. *J Clin Invest*, 105(4):513–20, 2000.

[20] J. Gai, M. Ji, and C. Shi et al. FoxO regulates expression of ABCA6, an intracellular ATP-binding-cassette transporter responsive to cholesterol. *Int J Biochem Cell Biol*, 45(11):2651–9, 2013.

[21] J. J. Wenzel, W. E. Kaminski, and A. Piehler et al. ABCA10, a novel cholesterol-regulated ABCA6-like ABC transporter. *Biochem Biophys Res Commun*, 306(4):1089–98, 2003.

[22] S. Honzumi, A. Shima, and A. Hiroshima et al. LXRalpha regulates human CETP expression in vitro and in transgenic mice. *Atherosclerosis*, 212(1):139–45, 2010.