Dissertation

submitted to the

Combined Faculties for the Natural Sciences and for Mathematics

of the Ruperto-Carola University of Heidelberg, Germany

for the degree of

Doctor of Natural Sciences

presented by

MSc Anthi Trasta

Born in: Athens, Greece

Oral-examination: 2nd of June 2017

Unraveling Gene Networks of Human Cholesterol Homeostasis and their Roles in Cardiovascular Disease

Referees:

Prof. Dr. Marko Kaksonen

Prof. Dr. Walter Nickel

SUMMARY

It is well established that elevated plasma cholesterol is a major risk factor for atherosclerosis, the cause of cardiovascular disease (CVD), which is most often manifested as coronary artery disease (CAD) and myocardial infarction (MI). The identification through GWAS of genes associated with lipids and CVD has only to a minor extent explained the genetic architecture of CVD and dyslipidemias. In an effort to identify genes affecting cholesterol regulation, an RNAi-based functional profiling of GWAS-derived loci, which were associated with lipid traits, CAD and/or MI was previously performed in our lab (Blattmann et al. 2013). This study resulted in the identification of 55 genes that had an effect on LDL internalization and/ or cellular cholesterol levels. However, most of the screen hits did not have a strong effect, suggesting that the combinatorial -rather than the individual- function of these genes might regulate cholesterol homeostasis and subsequently CVD. This reasoning is supported by the fact that CVD is a complex disease, which is assumed to arise from the synergistic effect of genes.

In the present study, a combinatorial RNAi screen was performed in order to identify interactions between genes identified in the aforementioned study, after juxtaposing them with the results of an Exome Chip of more than 70,000 individuals, genotyped for lipid traits (LDL, HDL, TG, TC) (Peloso et al. 2014). For this purpose, the effect on LDL uptake of all pairwise combinations between 30 candidate genes was tested, and 21 pairs were confirmed as genetic interactors. A gene interaction model network was constructed, based on the results of the screen, connecting known cholesterol regulators, as well as genes without a previously reported lipid-regulatory function. Secondary screens were performed to measure the effect of the gene interactions on *LDLR* mRNA and protein, as well as on *SREBF1* and *SREBF2* mRNA levels. The results from secondary experiments provided further valuable information for the mechanistic interpretation of the interactions. The one occurring between *LDLR*, which encodes for the receptor of LDL and *HAVCR1*, which encodes for a membrane receptor for hepatitis A virus was followed up. Furthermore, hypotheses were generated for the sub-network of *LDLR-MLXIPL-HAVCR1* on the mechanism of interaction influencing cholesterol homeostasis. Hypotheses were

made also for a few other interesting interactions, which correlated with cellular *LDLR* mRNA and/or protein levels, as well as with *SREBF* mRNA levels. For *HAVCR1*, mutation screening was performed, whereby the overexpression of 18 out of 19 mutations had a significant inhibitory effect on LDL uptake, further supporting a so far undescribed role for HAVCR1 in cholesterol endocytosis. In parallel, in collaboration with Heiko Runz (Merck Research Laboratories), all lead SNPs of the genes tested with co-RNAi were examined for co-occurrence and SNP-SNP interactions in a cohort of more than 4000 individuals (Muntendam et al. 2010). With this analysis, an additive effect was demonstrated for three pairs of SNPs that corresponded to gene interactions identified with the co-RNAi screen (*LPL*+*CELSR2*, *APOB*+*HMGCR*, *LDLR*+*NCAN*).

In summary, the study in hand identified combinatorial effects of genes on cholesterol homeostasis, through systematic identification of genetic interactions between GWAS-derived genes. Altogether, this research demonstrates the potential of the scalable strategy employed using quantitative cell-based assays, to uncover the genetic networks underlying common disorders and diseases. Further characterization of these networks would lead to a better understanding of CVD inheritance and provide valuable insight for the generation of novel treatments.

ZUSAMMENFASSUNG

Es ist gut bekannt, dass erhöhte Plasma-Cholesterinspiegel ein wichtiger Risikofaktor für Atherosklerose sind, die Ursache für Herz-Kreislauf (HKL)-Erkrankungen, die sich am häufigsten als koronare Herzkrankheit (KHK) und Myokardinfarkt (MI) manifestieren. Die Identifizierung durch genomweite Assoziationstudien (GWAS) von Genen, die mit Lipiden und HKL assoziert sind, hat nur in geringem Masse die genetische Architektur von gemeinsamen Erkrankungen wie HKL und Dyslipidämien erklärt. In einer Bemühung, Gene zu identifizieren, die die Cholesterin Regulation beeinflüssen, wurde zuvor in unserem Labor eine RNA-Interferenz-basierte funktionelle Profilerstellung von GWAS-abgeleiteten Genorten, die verbunden mit Lipidmerkmalen, KHK und/ oder MI geworden sind, durchgeführt (Blattmann et al. 2013). Diese Studie führte zur Identifizierung von 55 Genen, die einem Einfluss auf die LDL-Cholesterin Internalisierung und/ oder den zellulären Cholesterinspiegel hatten. Allerdings hatte für die Mehrheit der RNAi-Screen Treffer eine Unterdrückung ihrer Expression keinen starken Einfluss auf die Endozytose von LDL, was darauf hindeutet, dass die kombinatorischen -und nicht die individuelle Wirkung- dieser Genen die Cholesterin-Homöostase und infolgedessen KHK regulieren könnte. Diese Argumentation wird durch die Tatsache unterstützt, dass HKL eine komplexe Erkrankung ist, die aus Wechselwirkungen zwischen Genen entsteht.

In der vorliegenden Studie wurde ein kombinatorischer RNAi-Screen durchgeführt, um die genetischen Wechselwirkungen zwischen den in der obigen Studie identifizierten Genen aufzudecken, nachdem sie mit den Ergebnissen eines Exome-Chips von mehr als 70,000 Individuen (Peloso et al. 2014), die für Lipidmerkmalen (LDL, HDL, TG, Gesamtcholesterin) genotypisiert waren, verglichen wurden. Zu diesem Zweck wurde die Wirkung auf die LDL-Cholesterin Aufnahme aller gepaarten Kombinationen zwischen 30 Kandidatengenen getestet, und 21 Paaren als “Interaktoren” wurden bestätigt. Es wurde ein Gen-Interaktionsmodell Netzwerk auf der Basis der gefundenen genetischen Wechselwirkungen aufgebaut, wobei sowohl bekannte Cholesterolregler als auch Gene ohne eine zuvor berichtete Lipidregulationsfunktion verknüpft wurden. Folgeuntersuchungen wurden

durchgeführt, um die Wirkung der ko-RNAi-Screen Treffer auf *LDLR* mRNA und Protein Niveaus sowie auf *SREBF1* und *SREBF2* mRNA Niveaus zu messen. Die Ergebnisse aus diesen Folgeuntersuchungen lieferten weitere wertvolle Informationen für die mechanistische Interpretation der Wechselwirkungen. Die Interaktion zwischen *LDLR*, das für den Rezeptor von LDL kodiert, und *HAVCR1*, das für einen Membranrezeptor für Hepatitis-A-Virus kodiert, wurde mehr im Detail untersucht. Ausserdem wurden Hypothesen für das Subnetz von *LDLR-MLXIPL-HAVCR1* auf dem Mechanismus der Interaktion, die Cholesterin-Homöostase beeinflussen, erzeugt. Hypothesen wurden auch für einige andere interessante Interaktionen aufgestellt, die in Wechselbeziehung mit zellulären LDLR mRNA und /oder Protein Niveaus, oder mit SREBF mRNA Niveaus waren. Für das Gen *HAVCR1*, konnte bei der Überexpression von 18 aus 19 Mutationen, die in Patienten vorkommen, eine signifikante Hemmwirkung auf die LDL-Aufnahme beweist werden, was die wichtige Rolle des *HAVCR1* Gens bei der Cholesterol-Endozytose weiter zeigte. Parallel dazu wurden in Zusammenarbeit mit Heiko Runz (Merck Forschung Labore, Boston USA) alle führenden SNPs der bei ko-RNAi getesteten Genen auf gemeinsames Auftreten und SNP-SNP-Wechselwirkungen in einer Kohorte von mehr als 4000 Personen (Muntendam et al. 2010) getestet. Unter Verwendung dieser Analyse konnte ein additive Effekt bei drei Paaren von SNPs nachgewiesen werden, die den Gen-Wechselwirkungen, die beim ko-RNAi Screening identifiziert wurden, entsprachen (*LPL* + *CELSR2*, *APOB* + *HMGCR*, *LDLR* + *NCAN*).

Zusammenfassend hat die vorliegende Studie funktionellen Verbindungen zwischen bisher nicht verwandten Genen erstellt, die eine kombinatorische Wirkung auf die Cholesterin Homöostase nachweisen. Insgesamt zeigt die Studie das Potenzial dieser erweiterbaren Strategie zur systematischen Identifizierung von genetischen Wechselwirkungen zwischen GWAS-abgeleiteten Genen mit quantitativen zellbasierten Assays, um die genetischen Netzwerke, die Häufige Erkrankungen und Krankheiten zugrunde liegen, aufzudecken. Eine weitere Characterisierung dieser Netzwerke würde zu einem besseren Verständnis der HKL-Vererbung führen und einem wertvollen Einblick in die Entstehung neuartigen Behandlungen liefern.

ACKNOWLEDGEMENTS

First of all, I would like to thank Rainer Pepperkok, for giving me the opportunity to conduct my PhD in his group, as well as Heiko Runz, in whose group I have been working during the beginning of my PhD. I would like to thank them both for the supervision, critical discussions, suggestions, and assistance during my PhD. I would especially like to thank Rainer Pepperkok, for his support and for his always-positive feedback, and for helping me to overcome moments of disappointment with his broad perspective. I appreciate his direct and honest opinion, and I learned a lot from him, in terms of scientific thinking and writing. I would also like to thank the members of my Thesis Advisory Committee: Walter Nickel, Marko Kaksonen and Wolfgang Huber, for guidance during the yearly meetings. Special thanks go to Wolfgang Huber, for helpful discussions outside the TAC meetings. I would furthermore like to thank Prof. Dr. Marko Kaksonen and Prof. Dr. Walter Nickel, for being referees of this PhD thesis.

During my PhD, various people helped me and supported me. To begin with, I would like to thank the people that helped me in the beginning of my PhD; Brigitte Joggerst for teaching me the cellular assays, Christian Schuberth for guiding me through the imaging and analysis of functional assays, and Peter Blattmann for useful discussions. Following, I would like to thank the members of the Advanced Light Microscopy Facility -Christian Tischer and Volker Hilsenstein for helping me with image analysis, Beate Neumann and Sabine Reither for their assistance with array preparation and microscopy, Stefan Terjung and Yury Belyaev. Next, I would like to thank the GeneCore Facility (Vladimir Benes and especially Paul Collier) for assistance with the qPCR experiments. Special thanks go to Bernd Klaus from the Center for Statistical Data Analysis (Huber group) for his considerable assistance with statistical analysis.

Additionally, I would like to thank the Pepperkok team for making me feel like family during these years. Special thanks go to Magdalena Zimon, for continuous guidance, help, and useful discussions, as well as for critical reading of this manuscript. Her contribution to my PhD has been invaluable. Special thanks go also to Fatima Verissimo, for both scientific as well as moral support, especially during difficult

phases of the PhD. Many thanks go to Aliaksandr Halavatyi for his valuable help with coding and statistics, and for always being so ready to help and share his expertise. I would also like to thank Paolo, Sandra, Nurlanbek, Emiliana, Juan, as well as Muzamil and George for useful discussions and suggestions. Furthermore, I would like to thank Niki (Gayathri Vegesna) for support throughout this time and for the good times outside work. I would also like to thank Sanjana and Lottie for being so nice and supportive. Finally, I would like to thank Miriam Reiss for great help, technical support and useful discussions, and for being such a welcoming and friendly person. EMBL has been a great place to work and I enjoyed a lot its international, friendly and interactive atmosphere.

I would also like to thank all my friends in Heidelberg, for being part of this beautiful life chapter. Special thanks go to some very close family members, friends and mentors back in Greece, for their love and support during these years and for helping me to address challenges with patience, persistence and positive spirit.

Finally and most importantly, I would like to thank my family, without the support of which I would not have reached this point. I thank them for always being there for me, and for being proud of me. Special thanks goes to my mother, who has been motivating and encouraging me, even during hard times. This thesis is dedicated to her and to my beloved father, who sadly didn't stay in this world long enough to see his daughter become a Dr.

Table of Contents

LIST OF FIGURES

LIST OF TABLES

PUBLICATIONS

Anthi Trasta, Heiko Runz and Rainer Pepperkok
Mapping of cholesterol regulation networks through genetic interaction analysis by combinatorial RNAi.
(in preparation)

Conference posters and talks

Anthi Trasta, Heiko Runz and Rainer Pepperkok
"Unraveling Gene Interaction Networks of Human Cholesterol Homeostasis and their Role(s) in Cardiovascular Disease"
5th EMTRAIN PhD workshop: Innovative approaches to address unmet medical needs, Berlin, 24-27 April 2016
Poster & short talk

Anthi Trasta, Heiko Runz and Rainer Pepperkok
"Unraveling Gene Interaction Networks of Human Cholesterol Homeostasis and their Role(s) in Cardiovascular Disease"
FEBS Workshop: "Lipids as molecular switches", Spetses, Greece, 25-30 August 2014
Poster & short talk

Anthi Trasta, Heiko Runz and Rainer Pepperkok
"Unraveling Gene Interaction Networks of Human Cholesterol Homeostasis and their Role(s) in Cardiovascular Disease"
CBB Retreat May 2014, Abbey Mont Sainte Odile in France
Poster & short talk

1 INTRODUCTION

The present PhD Thesis endeavors to shed light on the genetic interactions that regulate cholesterol homeostasis in human, and are thus causally associated with the development of cardiovascular diseases (CVDs). With an aim to demonstrating the role of genes and their variants in CVD, this introduction starts with a general overview of disease genetics (1.1), which is followed by a more focused description of the genetics and pathophysiology of CVD (1.2). The causative link between cholesterol and CVD is explained (1.2.2) and the mechanisms of cholesterol regulation are illustrated thereafter (1.3). Given that the study in hand focuses on the role of genetic interactions in CVD, which is a complex disease, the association between genetic interactions and complex diseases is analyzed (1.4). In the last section of the introduction, the theoretical framework is provided for the approaches utilized in this study in order to screen for gene-gene interactions, through combinatorial perturbations of gene function with RNAi (1.5).

1.1 INVESTIGATING DISEASE GENETICS

As the potential to control infectious and nutritional diseases in developed countries increases, it becomes evident that genetic diseases are a major cause of disability, mortality and morbidity. Understanding the genetic basis of hereditary diseases is challenging, due to the diversity of the human genome, as well as the multiple genetic causative elements that might contribute to disease risk. The sequencing of the complete human genome (International Human Genome Sequencing Consortium 2004), together with a comprehensive description of common human genetic variation (The 1000 Genomes Project Consortium 2015) provided us with a plethora of genetic information, that is critical for the identification of disease-causing genes.

1.1.1 INHERITANCE OF DISEASE GENETICS

Genetic disorders arise from the malfunction of genes, most often caused by mutations. Genetic diseases caused by mutations in a single gene or locus are

described as “monogenic” or “Mendelian”. Examples of such single-gene diseases are Huntington’s disease and cystic fibrosis. However, a plethora of genetic disorders are caused by mutations in more than one gene and their inheritance cannot be explained by simple Mendelian theory; therefore, they are described as “non-Mendelian” or “polygenic”. The concept of “polygenic” inheritance was introduced as early as 1918 by R. Fisher, who analyzed the causes of human variability using mathematics to combine Mendelian genetics with natural selection. Fisher demonstrated that the individual phenotype of quantitative traits can be produced by the combined action of many different genes (Fisher 1918). Many diseases have been shown to have a complex, polygenic inheritance pattern. Common diseases, such as Alzheimer’s and Parkinson’s (Singleton and Hardy 2016), schizophrenia (Sawa and Snyder 2002), type 2 diabetes (Flannick and Florez 2016) and cardiovascular disease (Kathiresan and Srivastava 2012) are named “common complex diseases”, because they can be caused by interplay between many genes as well as between genes and environment. In order to investigate the causality of genetic diseases, one needs to delve into the genetic variation and acquire a thorough understanding of its contribution to the observed phenotype.

1.1.2 GENETIC VARIATION

The human genome is unique; this means that any two persons share 99.5% identity in their genomic sequences. Nonetheless, there are millions of differences among the 3.2 billion base pairs, which create the variation among individuals and may account for susceptibility to disease. Remarkably, the 1000 Genomes Project, in which 2,504 individuals from 26 human populations were sequenced, demonstrated that "a typical genome differs from the reference human genome at 4.1 million to 5.0 million sites … affecting 20 million bases of sequence" (The 1000 Genomes Project Consortium 2015). The vast majority of the variants are single-nucleotide polymorphisms (SNPs), namely substitutions of a single nucleotide at a particular site in the DNA sequence. The bulk of SNPs (~ 90%) are common SNPs, namely they have a minor allele frequency, MAF (MAF; the frequency at which the less frequent allele appears in a population) of ≥5%. Throughout the twenty-three chromosome pairs

there are roughly 11 million common SNPs. The remaining (~10%) SNPs are rare SNPs, namely they have a MAF less than 1%.

Common SNPs tend to be older than rare SNPs, which are recently derived alleles in the human population (Kruglyak and Nickerson 2001). In principle, these sequence variations have arisen from single historical mutation events, and are therefore associated with neighboring variants that were present in the ancestral chromosome on which the mutation occurred (Kruglyak and Nickerson 2001). Hence, a specific group of SNP alleles are found on a single chromosome that commonly segregate together, forming the so-called "haplotype blocks" or "haploblocks". The closest the distance between two SNPs on a chromosome, the higher the chance that they are inherited together, due to decreased recombination likelihood between them; this is described as "linkage disequilibrium" (LD). Within each haploblock, a SNP in high LD with the other alleles, termed "tag SNP" represents the region and can be used to identify the whole haplotype. It has been proved that genotyping a set of carefully selected "tag SNPs" is sufficient to uncover most of the genetic variation (The International HapMap Consortium 2003). This was confirmed by the International HapMap Project, which genotyped 3.1 million SNPs, in an effort to create a haplotype map of the human genome (The International HapMap Consortium 2007). Sequence data collected by the project demonstrated that SNPs are typically perfectly correlated with flanking SNPs and haplotypes; SNPs that show a strong correlation with other SNPs are termed "proxies" (The International HapMap Consortium 2005). Thus, when two SNPs are perfectly correlated (r^2=1.0, r^2; correlation coefficient), testing one of the two SNPs is sufficient to extract information for the effect of the other one.

1.1.3 Discovering Disease-Associated Variants

The pinpointing of causal genes and variants in order to uncover the genetic architecture of hereditary traits and diseases is done with genetic mapping (Altshuler, Daly, and Lander 2009). Genetic mapping is a powerful approach through which the relative position of causal genes on a chromosome can be determined, based on their inheritance pattern. In order to produce a genetic map, DNA has to be isolated from tissue or blood samples collected from patients, often members of

the same family in which a certain disease “runs”. Characteristic DNA patterns that are observed uniquely in affected individuals, called “markers” are useful to locate and identify the causal gene(s), based on LD.

1.1.3.1 Linkage Studies for Mendelian Disorders

The simplest form of genetic mapping, described as “genome-wide linkage analysis”, or simply “linkage analysis” was proposed in 1980, and represented the first approach to connect DNA sequence polymorphisms with Mendelian traits (Botstein et al. 1980). Linkage analysis has been used to investigate the inheritance of rare Mendelian diseases within families with many affected individuals, as the segregation of the disease phenotype is a prerequisite for linkage mapping (Botstein and Risch 2003). In linkage studies, a disease-associated chromosomal region (locus) is identified, by discovering the alleles- “markers” that segregate (are linked) with the disease, and are therefore assumed to neighbor with the disease-causing gene(s) (Mayeux 2005). Once the candidate locus is defined, the likelihood that the genetic marker and the disease gene are linked is statistically estimated, based on the genetic recombination frequencies generated by meiotic cross-over. In order to further narrow down the genomic region, where the disease gene must reside, LD is often used in a method termed “linkage disequilibrium mapping” or “LD mapping” (Botstein and Risch 2003). Thereafter, a strategy referred to as “positional cloning” is used to identify the causal mutation(s) and define its functional role in the pathogenesis of the disease (Puliti et al. 2007).

Classical linkage analysis and positional cloning have been successful in the identification of rare genetic mutations causing Mendelian traits and diseases, as well as Mendelian subtypes of common diseases. These methods failed, however, to unveil the causal genes in common forms of complex diseases.

1.1.3.2 Genome-Wide Association Studies for Complex Diseases

An alternative to family-based linkage study for the genetic mapping of complex diseases is the population-based “genetic association study”. The basic idea behind genetic association studies is to correlate genetic variants with a trait or disease, by comparing SNP frequencies between a sample of cases and a sample of control

individuals. The measure of association between each individual SNP and the disease is given by the odds ratio (OR), which reflects the probability of disease occurrence divided by the probability of non-occurrence, compared between cases and controls. In a “genome-wide association study” (GWAS), also known as “whole-genome association study” (WGAS), a genome-wide set of SNPs is examined in a large population (hundreds or thousands) of both affected and non-affected individuals, to detect any existing association(s) with a trait or disease (see Fig. 1). Owing to LD, which is widespread in the genome, a set of tag SNPs is sufficient to be genotyped, as it can capture most of the allelic variation in their region.

GWA studies of common variants (with a MAF≥5%), also known as “common-variant association studies”, CVAS, were based on the common disease-common variant (CDCV) hypothesis. The CDCV hypothesis postulates that susceptibility to common diseases is often attributable to a small number of common polymorphisms with low penetrance, namely at each disease-associated locus there is only one or a few predisposing alleles (Reich and Lander 2001; Schork et al. 2010) (see Fig. 2). Because of strong associations among SNPs, only a set of the best tag SNPs need to be genotyped in a GWAS to predict the remaining associated SNPs (see 1.1.2). The selection of SNPs is done based on LD patterns among them, making use of the HapMap that was built for this purpose; to facilitate the identification of common causal variants (The International HapMap Consortium 2005).

The first GWAS was published in 2005 and investigated age-related macular degeneration (Haines et al. 2005). Up to this point (03/2017), 2785 GWA studies have been published and 32284 unique SNP-trait associations have been identified (http://www.ebi.ac.uk/gwas/). GWA studies have been quite successful in identifying associations between common SNPs and disease traits. However, the majority of SNPs identified have a low effect size (strength of association) on the traits with which they are associated (Manolio 2009; Wang et al. 2005). To date, GWA studies have identified hundreds of genetic variants associated with complex human diseases, but most variants identified confer relatively small effects in risk (odds ratio (OR); 1.2-1.5), and explain only a small proportion of the population variance in liability to disease (typically 0.2-0.5%) (see Tbl. 1). Even in aggregate, all the risk SNPs identified so far for any complex disease explain 5-10% of variance, when the overall

heritability of these diseases is 40-80% (Manolio 2009; Zeggini and Morris 2010). Thus, there is a large proportion of the heritable component of complex diseases that remains unidentified, which is described as “missing heritability” (Maher 2008; Manolio 2009).

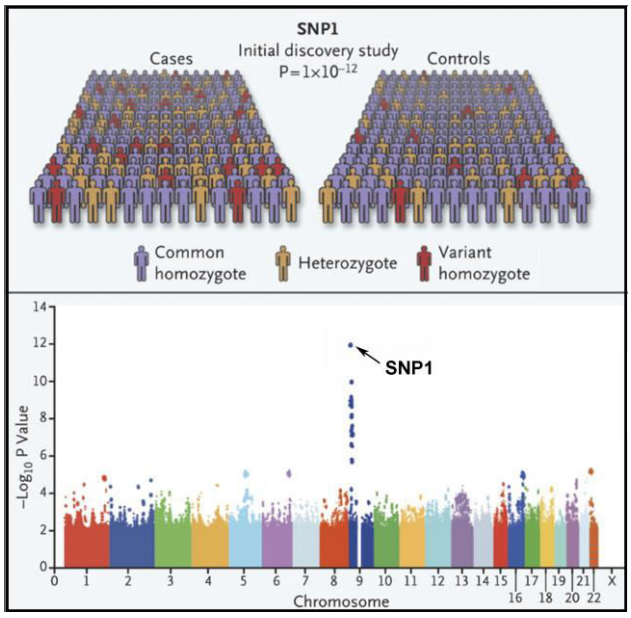


Figure 1. The Genome-Wide Association Study (GWAS)
In a typical GWA study, individuals genotyped for SNPs are divided in cases and controls. The strength of association between each SNP and disease is calculated based on the prevalence of the SNP in the two populations. In the example shown here, SNP1 is associated with the disease with P value of $1x10^{-12}$. The plot shows the P values for all the SNPs genotyped, located in different chromosomes.
(Reproduced with permission from Manolio NEJM 2010, Copyright Massachusetts Medical Society)

1.1.3.3 THE ISSUE OF MISSING HERITABILITY

The issue of missing heritability has been heavily discussed and many explanations for this have been suggested. One possible explanation is that only a small portion of the causal common variants have been identified so far, and the missing heritability is hidden within the ones that still need to be found -as many as 30% of existing common variants might remain undetected (Manolio 2009; Wang et al. 2005). A second explanation takes into account the rare variants (MAF<5%), which are assumed to have a larger effect than the common ones, but fail to be detected by GWASs due to their poor representation in genotyping arrays (Pritchard 2001) (see Fig. 2). Moreover, a probable cause for the small effect sizes observed so far is that the potential causal variants within each gene have been incompletely examined. That is to say, genes identified as having common associated variants could as well harbor rare variants with larger effects. This is implied by the identification of high-frequency lipid-associated variants in genes such as LDLR and PCSK9 that also carry low-frequency variants causing Mendelian dyslipidemias (Cohen et al. 2005; Kathiresan et al. 2008; Manolio 2009). Structural variants have also been suggested to account for missing heritability, as a special case of rare variants, based on findings for schizophrenia, autism, epilepsy and mental retardation (Cirulli and Goldstein 2010). Epigenetic effects such as genomic imprinting and parent-of-origin effects, that is sex-specific associations between variants in parents and risks in offspring, have also been implicated (Eichler 2010). Last but not least, gene-gene and gene-environment interactions, which cannot be captured by GWA studies, are believed to be the explanation for the missing inherited risk (Manolio 2009; McCarthy and Hirschhorn 2008).

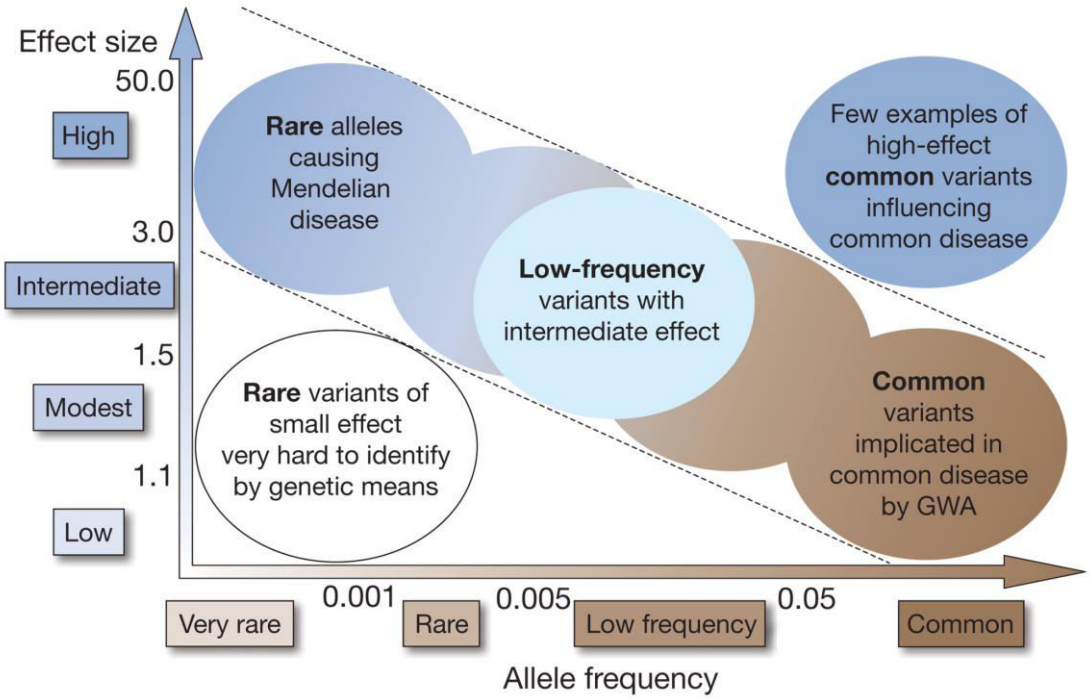


Figure 2. Feasibility of identifying genetic variants by risk allele frequency and strength of genetic effect (odds ratio)

Variants of low minor allele frequency (MAF), roughly 0.5%≤MAF≤5%, or rare (MAF<0.5%) are not frequent enough to be captured by GWA genotyping arrays, and usually do not carry large effect sizes to be detected by linkage analysis. Low frequency variants could however have considerable effect sizes (increasing disease risk two-to threefold) and could considerably contribute to missing heritability.

(Reprinted by permission from Macmillan Publishers Ltd: Nature, Manolio et al., copyright 2009)

Disease	Number of loci	Proportion of heritability explained	Heritability measure
Age-related macular degeneration	5	50%	Sibling recurrence risk
Crohn's disease	32	20%	Genetic risk (liability)
Systemic lupus erythematosus	6	15%	Sibling recurrence risk
Type 2 diabetes	18	6%	Sibling recurrence risk
HDL cholesterol	7	5.2%	Residual*phenotypic variance
Height	40	5%	Phenotypic variance
Early onset myocardial infarction	9	2.8%	Phenotypic variance
Fasting glucose	4	1.5%	Phenotypic variance

*Residual is after adjustment for age, gender, diabetes.

Table 1. Estimates of heritability and number of loci for several complex traits

For very few common diseases, such as age-related macular degeneration, a small number of common variants of large effect explain a substantial proportion of heritability. For most common diseases however, the proportion of heritability explained is minor, despite a much larger number of identified variants.

(Reprinted by permission from Macmillan Publishers Ltd: Nature, Manolio et al., copyright 2009)

1.1.3.4 EXOME SEQUENCING STUDIES FOR COMPLEX DISEASES

The extreme alternative to CDCV is the "heterogeneity hypothesis", also known as the "multiple rare-variant hypothesis" or "common disease, rare variant (CDRV) hypothesis". The CDRV hypothesis postulates that disease risk is attributed to multiple low-frequency variations (MAF<5%) at different loci, each with relatively high penetrance (Pritchard 2001; Schork et al. 2010; Smith and Lusis 2002). The argument for an important role of rare variants is that purifying selection purges deleterious alleles strongly predisposing to disease, which are thus kept at low frequencies (Goldstein et al. 2013; Pritchard 2001; Zuk et al. 2014). In order to sufficiently test the CDRV hypothesis against the CDCV hypothesis for any disease, the first step is to identify rare variations among affected individuals. This can be only achieved through DNA sequencing, and consitutes the target of "rare-variant association studies", often called "exome sequencing" studies.

Rare-variant association studies (RVAS), relative to common-variant association studies, test for association of rare variants with a trait or disease. RVAS, like GWAS require very large sample sizes. Unlike GWAS, wherein all individuals are genotyped, RVAS involve sequencing of all individuals, which may or may not be followed by genotyping. A key difference between RVAS and CVAS is that most RVAS have focused on the coding part of the genome -the exome, which constitutes only about 1.5% of the genome. This way the sequencing cost can be lowered for a large sample of patients, because less sequencing is required per individual. Another difference is that because the MAF of rare variants does not allow for association tests, RVAS test combined sets of rare variants instead of individual variants, in order to reach sufficient statistical power (Kiezun et al. 2013). In RVA studies, variants identified by aligning sequence reads to the reference genome, are annotated using probabilistic tools for their predicted functional effects (Do, Kathiresan, and Abecasis 2012).

Exome sequencing has been so far useful in studies of Mendelian disorders and is very promising for the study of complex traits. Ideally, exome sequencing can be employed to design association studies that will study both common and rare variants, as the genetic etiology of complex diseases is believed to involve both components. Noteworthy, the prevailing view for the allelic architecture of complex diseases falls between the two extreme theories that were aforedescribed -the CDCV

and the CDRV (Hall, Moore, and Ritchie 2016; Wang et al. 2005). Namely, there is evidence that both low- and high-frequency alleles account for disease risk.

1.1.3.5 eQTL studies for Complex Diseases

GWASs together with exome sequencing studies have brought to light a vast collection of genomic loci and variants -both common and rare- associated with common complex diseases. The challenge however is to acquire a mechanistic understanding of how these loci and variants affect diseases. Interestingly, it has been observed that most variants identified by GWASs reside in non-coding genomic regions, which suggests that they might exert their effect rather indirectly, through regulation of gene expression (Albert and Kruglyak 2015). A particular class of variants, named "expression quantitative trait loci (eQTLs)", includes genomic regions harboring variants that influence gene expression. Mapping of eQTLs is based on association analyses comparing the expression levels of genes between two sets of individuals, who have been grouped based on which allele of the variant under study they carry. eQTLs may exert their effect to a gene located either in a close distance (in-cis eQTLs) or far away (in-trans eQTLs) from them. Notably, it has been demonstrated that heritable gene expression regulation is at large due to variants functioning from a distance (Grundberg et al. 2012). This means that a SNP discovered within the defined region of a specific gene might have a functional role for a different gene, instead for the gene to which it territorially belongs. eQTL studies have been proven valuable in providing insight into the function and the causality of variants associated to complex traits and diseases.

A notable example is a SNP (rs12740374) at the 1p13 locus, which (locus) was reported by GWASs as associated with serum low-density lipoprotein cholesterol (LDL-c) and myocardial infarction (Samani et al. 2007). This SNP, which resides at the 3' untranslated region (3' UTR) of *CELSR2* gene, was shown to be part of a binding site for a transcription factor named C/EBP which regulates the transcription of *SORT1* gene (Linsel-Nitschke, Samani, and Schunkert 2010; Musunuru et al. 2011). Binding of C/EBP at the site of that SNP results to elevated expression of sortilin 1, which in turn leads to reduced LDL-c levels and subsequently to a lower risk for

myocardial infarction (MI), which is a cardiovascular event (see 1.2.2) (Kathiresan et al. 2009; Strong et al. 2012).

1.2 Cardiovascular Disease

Cardiovascular disease (CVD) is a leading health problem in the developed countries of the world, as well as many of the developing countries. CVD has been the leading cause of mortality since 1921, according to data from the Centers for Disease Control and Prevention (CDC) (https://www.cdc.gov/nchs/fastats/heart-disease.htm). According to CDC, an estimated 17.5 million people deceased from CVD in 2012, representing 31% of global deaths. Moreover, in the United States alone more than 80 million individuals are affected by the disease.

Cardiovascular disease comprises a number of conditions, ranging from myocardial infarction (MI) to congenital heart disease, most of which are heritable (Kathiresan and Srivastava 2012).

1.2.1 Genetic Architecture of Cardiovascular Disease

Some forms of CVD, such as familial hypercholesterolemia (FH) follow a Mendelian inheritance pattern, in which case a single causal gene has a large effect on the observed phenotype. In such subtypes of the disease, direct sequencing and family-based linkage analysis have proved successful in the identification of the causal genes. In the example of familial hypercholesterolemia, M. Lehrman from the laboratory of M. Brown and J. Goldstein identified the causative deletion in the gene of low-density lipoprotein receptor (LDLR) by directly sequencing the gene in an FH patient (Lehrman 1985).

However, most forms of CVD, such as MI or plasma lipid concentrations exhibit complex inheritance, suggesting the combined contribution of multiple genes and non-genetic factors to the disease. Complex forms of CVD are investigated with genetic association studies, which have focused until recently on common variants. The combination of GWAS with re-sequencing of selected genes has led to the conclusion that the genetic architecture of complex CVD traits is a mosaic of common variants with a small effect, rare variants with a large effect and environmental factors (Kathiresan and Srivastava 2012). A mosaic model hypothesis

for cardiovascular traits is reinforced by the strong overlap between genes identified with GWAS and those identified earlier by familial linkage analysis (Kathiresan and Srivastava 2012). In addition, it is believed that interactions among genetic and non-genetic factors shape the complex, multifactorial CVDs (Farhan and Hegele 2013).

In particular, complex CVDs are assumed to result from multiple inherited DNA variants, each of which has a small effect on disease risk, but which cumulatively could play a substantial role (Farhan and Hegele 2013; Lanktree and Hegele 2009). In the case of CAD, rare variants in at least nine genes have been identified by sequencing studies, which contribute to disease risk (Khera and Kathiresan 2017) (see Tbl. 2). In addition, GWA studies have identified 202 independent variants in 109 loci as associated with disease risk. All the identified variants together explain 28% of the total estimated heritability of CAD, leaving the remaining, so-called "missing heritability" unexplained (McPherson and Tybjaerg-Hansen 2016). As was already mentioned (see 1.1.3.2.1), one explanation for the "missing heritability" of CVD and other complex diseases is believed to lie within interactions among causative genes and their polymorphisms (Lanktree and Hegele 2009). That is to say, the genetic architecture of CAD is assumed to result from the combined action of alleles in many genes (Badano and Katsanis 2002). The reason why the studies conducted so far have not been able to address the matter of "missing heritability" of CVD is because they examined the effect of genetic factors in isolation, not allowing for potential interactions with other genetic factors (Cordell 2009).

1.2.2 Pathogenesis of Cardiovascular Disease

The most commonly observed form of CVD is coronary artery disease (CAD), which affects the coronary arteries of the heart. The process underlying CAD, known as atherosclerosis, starts with the accumulation of lipoproteins in the inner walls of blood arteries, as a consequence of elevated LDL-cholesterol (LDL-c) levels in the bloodstream (Lusis 2012). Large arteries are bound on their luminal side by an endothelium, which insulates the vessel wall from the blood. The accumulation of plasma lipoproteins -particularly low-density lipoprotein (LDL)- at the sub-endothelial region results to an inflammatory response, which involves the recruitment of monocyte-derived macrophages and lymphocytes. The macrophages

take up lipoproteins to form the cholesterol-engorged “foam cells”, some of which eventually die resulting to a lipid-rich necrotic core. The accumulation of cholesterol, cellular debris and smooth muscle cells that migrate to the region results in the formation of an atherosclerotic plaque, which decreases the blood flow to the heart (see Fig. 3). This condition is described as CAD, with its most common manifestation being myocardial infarction (MI), which results from the rapture of the plaque, and leads to the formation of a thrombus that blocks the blood flow to the heart (Lusis 2012). Respectively, when a cerebral artery becomes occluded by a thrombus, the blood flow to the brain is inhibited and this causes an ischemic stroke.

Many genetic risk factors have been identified by epidemiological studies for atherosclerosis development, including elevated levels of low-density (LDL) and very-low-density lipoprotein (VLDL), low levels of high-density lipoprotein (HDL), high levels of lipoprotein a, family history, obesity, hypertension and diabetes. Environmental factors such as smoking, high-fat diet and lack of exercise may also contribute to the disease. However, elevated serum cholesterol levels are probably the only risk factor that is sufficient to cause atherosclerosis (Glass and Witztum 2001).

Blood cholesterol levels, particularly those of LDL-c, have long been associated with a risk for CAD (D’Agostino et al. 2008; Keys 1966), and the work of Brown and Goldstein established a causal relationship between them (Brown and Goldstein 1974, 1986; Goldstein and Brown 1977). The effect of a class of cholesterol-lowering drugs named “statins” on cardiovascular mortality reduction in patients with hypercholesterolemia is a confirmation that high cholesterol is causally related to CVD (Gould et al. 1998).

Gene	Carrier frequency	Intermediate phenotype	CAD risk	Therapy to mimic protective variants
Inactivating mutations confer increased risk				
LDLR	1 in 221 (0.5%)	↑ LDL cholesterol	↑ 320%	Not applicable
LPL	1 in 249 (0.4%)	↑ Triglyceride-rich lipoproteins	↑ 84%	Not applicable
APOA5	1 in 216 (0.5%)	↑ Triglyceride-rich lipoproteins	↑ 120%	Not applicable
Inactivating mutations confer decreased risk				
PCSK9	1 in 50 (2%)*	↓ LDL cholesterol	↓ 88%	Alirocumab, evolucumab (approved by the FDA and EMA)
NPC1L1	1 in 650 (0.2%)	↓ LDL cholesterol	↓ 53%	Ezetimibe (approved by the FDA and EMA)
ASGR1	1 in 120 (0.8%)	↓ LDL cholesterol ↓ Triglyceride-rich lipoproteins	↓ 34%	None
APOC3	1 in 150 (0.7%)	↓ Triglyceride-rich lipoproteins	↓ 40%	Volanesorsen (formerly known as ISIS-APOCIII$_{Rx}$; phase III trials)
ANGPTL4	1 in 360 (0.3%)	↓ Triglyceride-rich lipoproteins	↓ 53%	REGN1001 (preclinical development)
LPA	1 in 285 (0.4%)	↓ Lipoprotein(a)	↓ 24%	AKCEA-APO(a)-L$_{Rx}$ (phase II trials)

Table 2. Summary of results from gene sequencing studies for CAD

Damaging mutations in at least nine genes have been robustly associated with risk of coronary artery disease; in each case, identified genes disrupt pathways related to low-density lipoprotein (LDL) cholesterol, triglyceride-rich lipoproteins or lipoprotein (a) metabolism. Pharmacological therapies are in current use or development to mimic the protective variants for five of the six genes in which inhibition of the related protein would be predicted to reduce risk.

(Adapted by permission from Macmillan Publishers Ltd: Nature Rev. Genet., Genetics of coronary artery disease: discovery, biology and clinical translation, Khera and Kathiresan, copyright 2017)

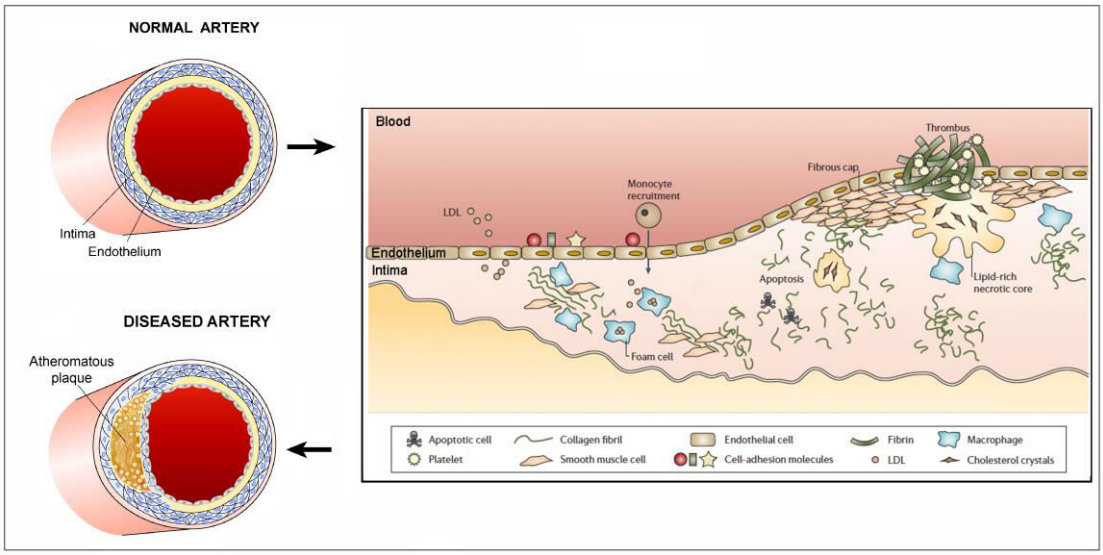


Figure 3. The process of atherosclerotic plaque development

Low-density lipoprotein (LDL) enters the arterial endothelium and induces an inflammatory reaction, which attracts monocytes and macrophages. Foam cells (lipid-loaded macrophages) accumulate and smooth muscle cells proliferate, which results to the plaque formation. Many foam cells and muscle cells die through apoptosis, making the plaque vulnerable. Plaque rupture causes formation of a thrombus, which can occlude the vessel.

(Adapted by permission from Macmillan Publishers Ltd: Nature Watkins and Farrall, copyright 2006 and reprinted from Trends Genet. 28 (6): 267-275, Genetics of atherosclerosis, Lusis, copyright 2012, with permission from Elsevier.)

1.3 CHOLESTEROL HOMEOSTASIS

Cholesterol is an essential component of mammalian cells, as it imparts special biophysical properties to their membranes, which can thus maintain their stability and durability. Furthermore, cholesterol partakes in the compartmentalization of cells, as it contributes to the creation of a semi-permeable barrier between specialized functional spaces (Ikonen 2008). Moreover, it participates in intracellular transport, cell signaling and -specifically in neurons- nerve conduction. Additionally, cholesterol is a precursor molecule for the biosynthesis of steroid and sex hormones, as well as vitamin D (Harvey and Ferrier 2011). Both dietary intake and biosynthesis constitute cholesterol sources for the body and tissues. In order to maintain cholesterol in the blood at the right level, and avoid atherosclerosis and CVD, the biosynthesis and metabolism of cholesterol are tightly regulated both in the body and in the cells.

1.3.1 REGULATION OF CHOLESTEROL IN THE BODY

Because cholesterol is not soluble, it is carried by several lipoprotein particles, which are classified according to their density into five groups. These classes from least dense to most dense are chylomicrons, very low-density- (VLDL), intermediate-density- (IDL), low-density- (LDL) and high-density- (HDL) lipoproteins, with LDL being the most prevalent. The role of lipoproteins is to transfer lipids through the extracellular fluid and serum to the different tissues of the body. Liver plays a central role in cholesterol metabolism in human, by maintaining cholesterol balance in the body and keeping LDL cholesterol levels steady in the plasma. To begin with, dietary cholesterol, which is initially absorbed by enterocytes of the small intestine, is subsequently delivered to the liver within chylomicrons, which consist of cholesterol and triglycerides. Moreover, the liver itself synthesizes cholesterol, which it then secretes in the form of VLDL. As much as 70% of VLDL that is produced by the liver is reabsorbed by it, while the remaining 30% is converted to LDL and cleared by extrahepatic tissues. Additionally, the liver absorbs cholesterol that is produced by the extrahepatic tissues in the form of HDL (see Fig. 4). It has been demonstrated that under normal dietary conditions the extrahepatic tissues are responsible for >80% of total sterol synthesis (Dietschy, Turley, and Spady 1993).

Whereas in a fetus or a child, there needs to be a net accumulation of cholesterol required for growth, in adults the amount of cholesterol that is both synthesized and absorbed must equal the amount of cholesterol that is excreted (plus minor quantities used for the skin or hormone synthesis) (Dietschy et al. 1993). The hepatic parenchyma cells have a key role in cholesterol excretion, by secreting sterols into the bile, in the form of bile acid, which is then disposed of with the feces (Dietschy et al. 1993).

Cells of hepatic and peripheral tissues have a central role in regulating the circulating cholesterol levels, by taking up LDL. The mechanism of LDL uptake from cells via binding to its receptor on the plasma membrane, LDL receptor (LDLR), significantly contributes to the clearance of LDL-cholesterol from the bloodstream (Brown and Goldstein 1986). This is evidenced by the fact that in absence of functional LDL receptors, LDL is massively accumulated, as was demonstrated in patients with the homozygous form of FH (Goldstein and Brown 1977).

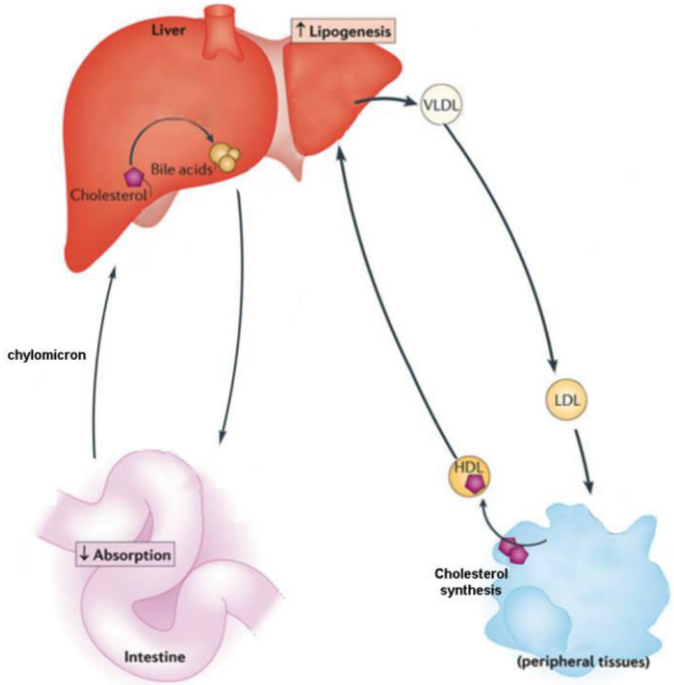


Figure 4. Cholesterol metabolism in the body.

Dietary cholesterol is initially absorbed by intestine, and subsequently delivered to the liver within chylomicrons. The liver synthesizes cholesterol, which it then secretes in the form of VLDL. 30% of VLDL that is produced by the liver is converted to LDL and cleared by extrahepatic tissues. Additionally, the liver absorbs cholesterol that is produced by the extrahepatic tissues in the form of HDL.

1.3.2 Regulation of Cholesterol inside the Cell

Within the cell, cholesterol is distributed unevenly among membranes, with a higher concentration being that of the plasma membrane, where it consists 20-25% of total lipids. In addition, cholesterol is ample in the endocytic recycling compartment and the Golgi complex (Ikonen 2008). Cholesterol sources for the cell consist of the following three major pathways; cholesterol synthesis from acetyl-coenzyme A (acetyl-coA), receptor-mediated LDL-cholesterol uptake, as well as hydrolysis of cholesterol esters stored inside lipid droplets and fatty acids.

1.3.2.1 Cholesterol Biosynthesis

Cholesterol biosynthesis is adaptable to the changing needs of the body for tissue growth and membrane remodeling. Moreover, *de novo* cholesterol biosynthesis is decreased when cholesterol intake increases. Importantly, the total sterol requirements of the body can be met by biosynthesis alone, even in total absence of dietary cholesterol (Dietschy et al. 1993).

All nucleated cells can synthesize cholesterol through the mevalonate pathway. The initial steps in the pathway result to the conversion of acetyl CoA to 3-hydroxy-3-methylglutaryl-coA (HMG-CoA), which is then converted to mevalonate by HMG-CoA reductase (HMG-CoAR or HMGCR) (Russell 1992) (see Fig. 5).

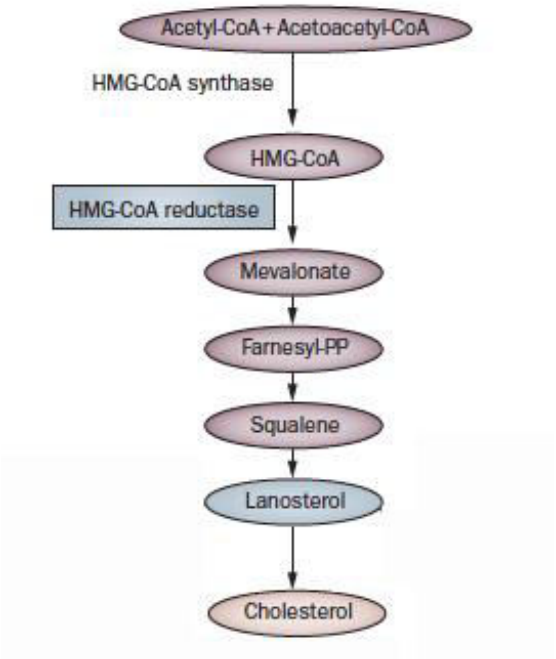


Figure 5. Cholesterol biosynthetic pathway
Shown are the basic steps in the cholesterol biosynthetic pathway from acetyl-coA and acetoacetyl-coA.
(Adapted by permission from Macmillan Publishers Ltd: Nature Rev. Karasinska and Hayden, copyright 2011)

HMGCR is regulated through a negative feedback loop mediated by sterols and other mevalonate metabolites and is considered to be the rate-limiting enzyme in the cholesterol biosynthetic pathway (Brown and Goldstein 1980). Importantly, HMGCR constitutes the target of statins, a class of cholesterol-lowering drugs that inhibit the enzyme by binding to its catalytic domain and thus blocking access of the natural substrate HMGCoA (Gelissen and McLachlan 2014).

Cholesterol synthesis is primarily controlled at the level of HMGCR, both transcriptionally and post-transcriptionally. The transcription of HMGCR is regulated by a class of transcription factors named "sterol regulatory element-binding proteins" (SREBPs) (Ikonen 2008). An additional level of regulation is achieved with the degradation of HMGCR, which is mediated by lanosterol, a cholesterol precursor. Lanosterol post-transcriptionally regulates HMGCR, by mediating its binding to INSIG (insulin-induced gene protein) in the ER, which triggers its ubiquitination and subsequently its degradation (Song, Javitt, and DeBose-Boyd 2005).

1.3.2.2 THE SCAP-SREBP AND THE LXR PATHWAY

The transcriptional regulation of cholesterol homeostasis is mediated by two main nuclear receptor systems; the first system comprises the aforementioned sterol regulatory element-binding proteins (SREBPs) while the second one consists of the liver X receptors (LXRs). The SREBP system induces the transcription of genes related to cholesterol synthesis (Goldstein, DeBose-Boyd, and Brown 2006), whereas the LXR system suppresses LDL-cholesterol uptake and promotes reverse cholesterol transport, namely cholesterol clearance from peripheral cells and excretion of bile acid (Zelcer 2009).

The regulation of cholesterol homeostasis by SREBP was elucidated by Brown and Goldstein and has been since extensively studied. Under conditions of ample sterol in the cytosol, SREBP is bound with high affinity to the chaperone SREBP cleavage-activating protein (SCAP), and the SCAP/SREBP complex is retained in the endoplasmic reticulum (ER) by the ER-anchor-protein; INSIG. Conversely, a decrease in cellular cholesterol level induces conformational changes to SCAP, which is then released from INSIG, thus allowing the SCAP/SREBP complex to be transported within COPII-coated vesicles to the Golgi, where SREBP undergoes proteolytic

processing (Goldstein et al. 2006). The cleavage product is a soluble fragment of SREBP, which enters the nucleus and binds to promoters containing the sterol regulatory element 1 (SRE-1) sequence, whereby it acts as mature transcription factor. SRE-1 is contained in genes involved in lipid biosynthesis and uptake, such as *HMGCR* and *LDLR*, and binding of SREBP to it leads to the upregulation of their expression and subsequently to increased cholesterol biosynthesis (Brown and Goldstein 1997; Ikonen 2008) (see Fig. 6).

In humans, the SREBP family consists of three members; SREBP1 has two isoforms, SREBP-1a and SREBP-1c, which are encoded from a single gene with alternative splicing, whereas SREBP2 is encoded by a separate gene. SREBP-1a promotes the synthesis of cholesterol and fatty acids, while SREBP2 activates the LDLR gene and enzymes of cholesterol biosynthesis (Goldstein and Brown 2015). SREBP-1c activates the synthesis of fatty acids, therefore its role in cholesterol regulation is not so relevant (Shimano et al. 1997).

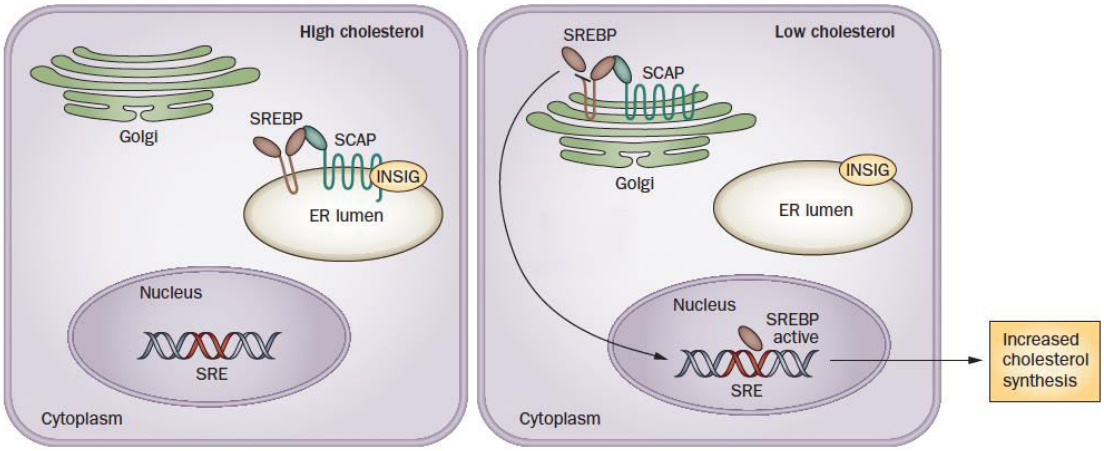


Figure 6. The SCAP-SREBP system

When cholesterol levels are high, SREBP is bound to SCAP, and the SREBP/SCAP complex is retained in the ER by INSIG. A decrease in cholesterol levels induces the release SCAP/SREBP from INSIG, allowing the complex to be transported to the Golgi, where SREBP undergoes proteolytic processing. The cleavage product is a soluble fragment of SREBP, which enters the nucleus and binds to promoters containing the sterol regulatory element 1 (SRE-1), whereby it acts as transcription factor for cholesterol regulating genes such as HMGCR and LDLR, and increases cholesterol biosynthesis.

Upon increase of intracellular cholesterol levels, the LXR pathway is also activated. This results to transcriptional induction of IDOL (Inducible degrader of LDLR)/MYLIP, an E3-ubiquitin ligase, that causes ubiquitylation of LDLR, which is thus targeted for lysosomal degradation (Nelson et al. 2015; Zelcer 2009). Subsequently, the amount of LDLRs on the membrane is reduced and the LDL uptake is inhibited. Therefore, the LXR/IDOL/LDLR axis serves as a complementary pathway to the INSIG/SCAP/SREBP pathway for regulation of cholesterol uptake.

1.3.2.3 CHOLESTEROL UPTAKE

Cholesterol is taken up by cells in the form of LDL, via a receptor-mediated pathway. LDL receptors (LDLRs) reside at clathrin-coated pits on the cell membrane, where they are anchored with their transmembrane domain (Brown, Herz, and Goldstein 1997; Goldstein et al. 1985). LDLRs bind particles that contain APOB or APOE apoproteins, such as chylomicrons, VLDL and LDL. When LDLRs bind to their ligand, LDL, the coated pits invaginate into the cell and pinch off the membrane to form clathrin-coated vesicles. After discarding their clathrin coats, the vesicles typically fuse to an early endosome, wherein LDL is released from its receptor. Subsequently, LDL is carried to lysosomes, where cholesterol esters are hydrolysed by acid lipase to release free cholesterol, while LDLR recycles back to the membrane (Lagor and Millar 2010) (see Fig. 7).

The cholesterol that is released from the endolysosomal system undergoes trafficking to other compartments, such as the plasma membrane, ER, recycling endosomes and mitochondria. Release of cholesterol from the late endosomes and lysosomes is strongly dependent on the NPC (Niemann-Pick type C) proteins; NPC1 and NPC2. Loss-of-function mutations in either of the two proteins results in a severe childhood lysosomal storage disease, named Niemann Pick type C (NPC). This disease is characterized by the accumulation of unesterified cholesterol and glycosphingolipids in the late endosomes and lysosomes (Blanchette-Mackie 2000; Ribeiro et al. 2001).

The LDLR which is released from the late endolysosomal compartment can alternatively be bound by PCSK9 (proprotein convertase subtilisin/kexin type 9), a protease that targets the receptor for degradation in the lysosomes. The function of

PCSK9 inhibits the recycling of LDLR from endosomes to the cell surface, leading to reduction of the amount of LDLRs available to internalize LDL, which causes an increase of LDL-cholesterol in the circulation (Lagor and Millar 2010). For this reason, PCSK9 has been the target of new cholesterol-lowering therapies that are based on its inhibition. Importantly, clinical trials with monoclonal antibodies against PCSK9 have demonstrated that PCSK9 inhibition can lower LDL-cholesterol levels by 60%-70% when administered either as monotherapy, or in combination with statins (Lepor and Kereiakes 2015). In accordance with its role in cholesterol metabolism, PCSK9 expression is also regulated by SREBP, through an SRE binding site on the promoter of PCSK9 gene (Lagor and Millar 2010).

Efficient internalization of LDLR in the liver requires the protein ARH (autosomal recessive hypercholesterolemia), also known as LDLRAP1 (LDLR adaptor protein 1), that binds to the cytoplasmic tail of LDLR (Lagor and Millar 2010). ARH was initially discovered in patients with homozygous ARH deficiency, which show a markedly decrease in LDL clearance from plasma (Lagor and Millar 2010). Last but not least, another adaptor protein, important for LDLR internalization is DAB2 (disabled homolog 2), which catalyzes the efficient clustering of LDLR into coated pits. DAB2, like LDLRAP1, mediates its effect via binding to the cytoplasmic tail of LDLR. In the absence of DAB2, LDLRAP1 can mediate the endocytosis of LDLR, only when AP-2 (Adaptor Protein complex 2) is present. DAB2 on the other hand can function independently of ARH and AP-2 (Maurer and Cooper 2006).

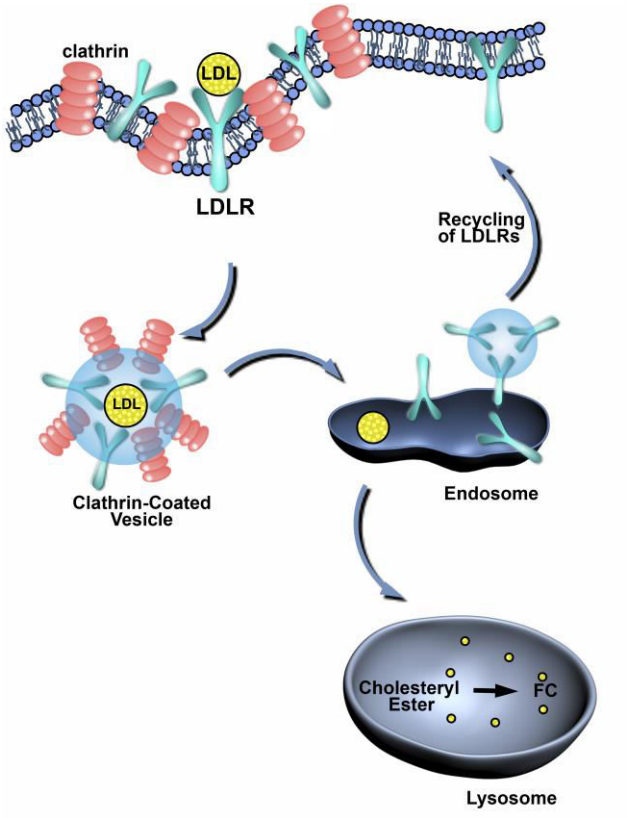


Figure 7. Cholesterol uptake

LDL-cholesterol is taken up from LDLR on clathrin-coated pits, which pinch off the membrane to form clathrin-coated vesicles. LDL is discharged from its receptor, LDLR, in endosomes, and cholesterol esters are hydrolyzed in lysosomes to release free cholesterol. LDLR is recycled from endosomes back to the plasma membrane.

1.3.2.4 CHOLESTEROL STORAGE

When intracellular cholesterol levels exceed the capacity of cellular membrane to incorporate sterols, cholesterol is either released from the cells or stored for future use. Excess free cholesterol in the cytosol is esterified by the enzyme acyl-coA: cholesterol acyltransferase-1 (ACAT-1), and cholesterol esters are stored within lipid droplets in the cytoplasm (Glass and Witztum 2001). Respectively, in peripheral cells excess cholesterol is converted to cholesterol esters by lecithin-cholesterol acyltransferase (LCAT). If needed, cholesterol esters within lipid droplets can in turn be hydrolyzed by hormone-sensitive lipase, to release free cholesterol (Glass and Witztum 2001). Cells rich in lipid droplets are mature adipocytes, as well as foam cells; the macrophages that engorge massive amounts of LDL-cholesterol during the atherosclerotic plaque development.

1.3.2.5 CHOLESTEROL REMOVAL

Cholesterol removal from macrophages is crucial for the prevention of atherosclerosis development. Macrophages secrete cholesterol in the form of HDL, which returns to the liver through reverse cholesterol transport and can from there be secreted to the bile. Essential regulators of cholesterol efflux are the ABC transporters. In particular, ABC transporter A1 (ABCA1) is the key regulator of HDL particle formation and secretion, and thus of maintenance of plasma HDL levels (Favari et al. 2015). ABCA1 binds the apoprotein APOA1 of HDL and facilitates cholesterol removal from late endosomes, in cooperation with the second ABC transporter, ABC transporter G1, (ABCG1). Hepatocytes internalize HDL through binding to the scavenger receptor SR-BI (Favari et al. 2015).

In enterocytes the main cholesterol transporter is ABCG5/G8, which mediates cholesterol efflux from the cells to the gut lumen. Hepatocytes on the other side secrete cholesterol and triglycerides in VLDL, the assembly of which is mediated by MTP (microsomal triglyceride transfer protein). The VLDL particles are carried within COPII-coated vesicles from the ER to the Golgi, where they are further lipidated before secretion (Ikonen 2008).

Endosomal cholesterol trafficking is additionally regulated by several Rab GTPases, including Rab8, Rab11, Rab7 and Rab9. Rab7 and Rab9 are involved in the removal of

cholesterol from late endosomes and lysosomes, whereas Rab8 and Rab11 function in recycling endosomes, and promote cholesterol efflux (Ikonen 2008).

1.3.3 GENETICS OF CHOLESTEROL REGULATION

So far, a large number of genes have been involved in cholesterol homeostasis and disorders in lipoprotein metabolism, which lead to the development of various diseases that are highly heritable. Numerous monogenic, mendelian disorders have been described which are caused by mutations in one of the genes regulating cholesterol synthesis, uptake, or efflux. The prototypic example is the aforementioned familial hypercholesterolemia (FH), also known as autosomal dominant hypercholesterolemia, which is mostly due to mutations in LDLR (Brown and Goldstein 1986). Additionally, mutations in two more genes - PCSK9 and APOB - were shown to be causally related with FH (Abifadel et al. 2003; Soria et al. 1989). The case of autosomal recessive hypercholesterolemia (ARH), which is caused by mutations in the LDLRAP1/ARH gene was also previously described (Garcia 2001). Furthermore, autosomal recessive Niemann-Pick type C disease is caused by mutations in NPC genes, as was already mentioned. Mutations in ABCG5 or ABCG8 cause sitosterolemia, a disorder whereby there is excessive accumulation of plant sterols (phytosterols) in tendons and arteries (Patel et al. 1998). Some additional examples of monogenic lipid disorders are shown in Table 3.

Despite the fact that various Mendelian lipoprotein disorders have been elucidated for each of the three lipoprotein traits -LDL, HDL and triglycerides- only a small proportion of the overall heritability of serum lipoprotein and lipid concentrations can be explained by them. Since the levels of plasma lipids follow a normal distribution in the general population, their inheritance pattern is believed to follow a rather polygenic model (Pirucello and Kathiresan 2010).

Gene	Protein	Locus	Associated traits	Lipid disorder
ABCA1	ATP-binding cassette transporter (ABC) A1	9q31.1	Low HDL-C	Tangier disease
ABCG5	ATP-binding cassette transporters (ABC) G5	2p21	High LDL-C	Sitosterolemia
ABCG8	ATP-binding cassette transporters (ABC) G8	2p21	High LDL-C	Sitosterolemia
APOA1	Apolipoprotein A1	11q23	Low HDL-C	ApoA-I deficiency
APOA5	Apolipoprotein A5	11q23	High VLDL, high chylomicrons	ApoA-V deficiency
APOB	Apolipoprotein B	2p24	Low LDL-C High LDL-C	Familial hypobetalipoproteinemia Familial defective ApoB-100
APOC2	Apolipoprotein C2	19q13	High chylomicrons	Familial ApoC-II deficiency
APOE	Apolipoprotein E	19q13	High VLDL, high chylomicrons	Familial dysbetalipoproteinemia
CETP	Cholesterol ester transfer protein	16q13	High HDL-C	Cholesteryl ester transfer protein deficiency
LCAT	Lecithin:cholesterol transfer protein	16q22	Low HDL-C	Lecithin-cholesterol acyltransferase deficiency (fish-eye disease)
LDLR	LDL receptor	19p13	High LDL-C	Familial hypercholesterolemia
LDLRAP1/ ARH	LDL receptor associated protein 1	1p36	High LDL-C	Autosomal recessive hypercholesterolemia
LIPA	Lysosomal acid lipase	10q23.31	Normal lipid levels	Wolman syndrome, cholesteryl ester storage disease
LIPC	Hepatic lipase	15q22	High VLDL remnants	Familial hepatic lipase deficiency
LMF1	Lipase maturation factor 1	16p13	High triglycerides	Combined lipase deficiency
LPL	Lipoprotein lipase	8p21	High chylomicrons	Lipoprotein lipase deficiency
MTTP	Microsomal triglyceride transfer protein	4q24	Low LDL-C	Abetalipoproteinemia
PCSK9	Proprotein subtilisin/kexin 9	1p32	Low LDL-C High LDL-C	PCSK9 deficiency Autosomal-dominant hypercholesterolemia
SAR1B	Saccharomyces cerevisiae homolog 1B	5q31.1	Low chylomicrons	Chylomicron retention disease

Table 3. Genes implicated in monogenic lipid disorders

Shown are the single gene causes for some primary dyslipidemias and their lipid phenotypes.

(Adapted from Teslovich 2010 and Kuivenhoven and Hegele 2014.)

1.4 GENETIC INTERACTIONS IN COMPLEX DISEASES

The identification through GWASs of genes associated with lipid traits and lipid-based disorders, has solved only a small part of the inheritance puzzle of such disorders, like dyslipidemia and CAD. The complete picture of the genetic architecture of these complex diseases still remains to be uncovered. Considering that complex diseases such as CAD are assumed to result from the combined action of alleles in many genes, interactions between the multiple identified genes should be examined (Badano and Katsanis 2002; Farhan and Hegele 2013; Lanktree and Hegele 2009). It is anticipated that detecting interactions between loci will both elucidate the biological and biochemical pathways that underlie complex traits and diseases, and contribute to a better understanding of the "missing heritability" (see 1.1.3.2.1).

Interestingly, it has been suggested that most of the genetic variants that contribute to the heritable phenotypic variation of complex diseases might have already been discovered, however the heritability explained by them was underestimated by not accounting for interactions among them. This hypothesis is evidenced by the example of Crohn's disease; the risk loci that have been identified so far by GWASs explain only 21.5% of the estimated heritability of the disease. However, if potential genetic interactions between the identified variants are accounted for, that is if non-additivity of effects is also considered, the explained heritability reaches 84% (Baryshnikova et al. 2013; Zuk et al. 2012).

1.4.1 GENETIC INTERACTIONS IN CORONARY ARTERY DISEASE

Regarding CAD, the total of 202 variants that have been so far identified with GWASs explain only 28% of the total estimated heritability, while the remaining "missing heritability" is believed to be lurking within interactions among polymorphisms of causative genes (Badano and Katsanis 2002; Lanktree and Hegele 2009; McPherson and Tybjaerg-Hansen 2016), as was aforementioned (see 1.1.3.2.1).

It is worth mentioning the following two examples of replicated gene-gene interactions in CAD, which were identified in candidate gene studies of common SNPs. The first example is the interaction between polymorphisms in the renin-angiotensin system and CAD. More specifically, an interaction between a variant in

the *ACE* (angiotensin I converting enzyme) gene and a variant in the *AGTR1* (angiotensin II type 1 receptor) gene was detected, that increases the risk for myocardial infarction (Kardia et al. 2006; Tsai et al. 2007). The second example is the interaction between *LPL* (lipoprotein lipase) gene and the ε4 allele of *APOE* (apolipoprotein E) gene that significantly increased the risk for a combined hypertriglyceridemia and hyperapobetalipoproteinemia "hyperTG/hyperapoB" phenotype, which is associated with elevated CAD risk (Perron et al. 2007). These two genes had also been demonstrated earlier to interact in HDL concentration levels (Corella et al. 2002). The above examples strongly support that genetic interactions should be taken into account when assessing the CAD risk.

1.4.2 THE CONCEPT OF STATISTICAL INTERACTION

The interaction between genes is termed "epistasis". The concept of epistasis was first invented by W. Bateson in 1909, to describe the deviations between the segregation ratios that were predicted based on a mendelian, single-gene inheritance model and the observed phenotypic outcomes (Phillips 1998). R. Fisher introduced "statistical epistasis" in 1918, in which the effect on a quantitative phenotype of two alleles at different loci deviates from additivity in a statistical model (Cordell 2002; Fisher 1918). In the case of complex quantitative traits, statistical epistasis refers to interaction between two or more loci resulting in a phenotype that cannot be predicted by simply adding the single-locus effects (Carlborg and Haley 2004) (see Fig. 8). Although epistasis was originally described by Fisher as a divergence from an additive model of gene action, population geneticists later on used a multiplicative model instead. Depending on the nature of the trait under investigation, either of two models -additive or multiplicative- is used nowadays to measure epistasis.

Most commonly, statistical interaction is described as departure from a linear (regression) model that describes the relationship between two (or more) predictor variables and a phenotypic outcome variable (Cordell 2009). Linear regression is a statistical approach based on the least squares method, commonly used for modeling quantitative traits, as well as for detecting gene-gene interactions in such traits. In particular, a linear model relating two non-interacting predictor variables

(X_1, X_2) to an outcome variable (Y) would look like $Y=\beta_0+\beta_1X_1+\beta_2X_2+\varepsilon$, where β_0 is the intercept, β_1 and β_2 are the slope or regression coefficients, and ε is the unexplained error of the model. On the other hand, if the two predictor variables (X_1, X_2) were interacting, the interaction term $\beta_3X_1X_2$ would be added to the linear model, which in this case would look like $Y=\beta_0+\beta_1X_1+\beta_2X_2+\beta_3X_1X_2+\varepsilon$. Thus, the inclusion of the product term $\beta_3X_1X_2$ to the linear model is used to measure deviations from additivity (Gilbert-Diamond and Moore 2011).

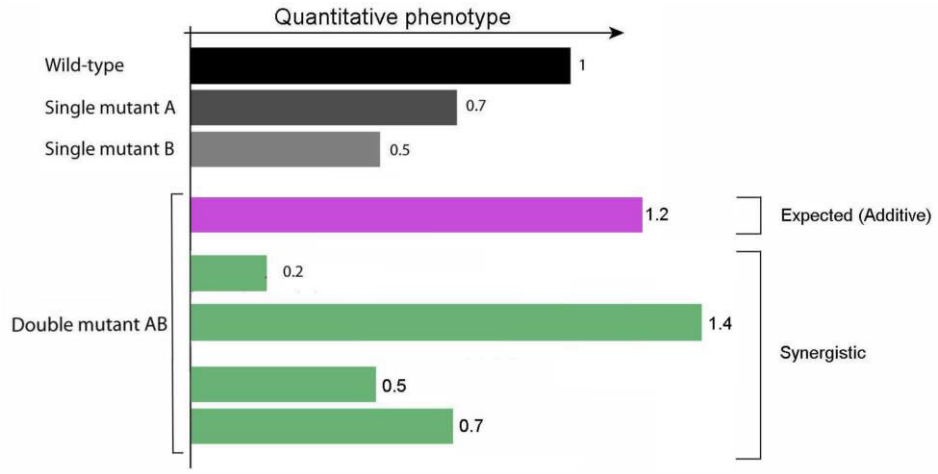


Figure 8. Graphical representation of genetic interaction inference from a measurable phenotype
The wild-type phenotype is defined as 1.0. The phenotypes of two mutants are 0.7 (single mutant A) and 0.5 (single mutant B). The expected phenotype of the AB double mutant based on an additive model would therefore be 1.2. Deviations from expectation are considered as interactions (synergistic phenotype).
(Adapted from Dixon et al., Annu. Rev. Genet. 2009)

1.4.3 Classification of Genetic Interactions

Based on the definition of epistasis, a genetic interaction is measured as the degree to which the resulting phenotype deviates from the additive (or multiplicative, depending on the trait under study) expectation. Namely, an additive phenotype is considered as a sign of absence of genetic interaction. Along these lines, genetic interactions can be classified as positive or alleviating, when the resulting phenotype is less severe than expected, and as negative or aggravating, when it is stronger than expected (Baryshnikova et al. 2013). Both positive and negative genetic interactions are considered synergistic, that is when the phenotype resulting from the combination of two variants differs from the expected phenotype of the individual variants (Phillips 2008).

When studying genetic interactions in double mutants (or double knockouts/ knockdowns), the aforedescribed classification into epistatic, suppressive or synergistic can be adopted. Specifically, a double mutant phenotype is considered epistatic when it resembles one of the single mutants, while the suppressive phenotype resembles the wild-type condition rather than any of the single mutants. Furthermore, a phenotype is described as synergistic, when the joint effect of two mutations is either stronger or milder than their individual effects (Pérez-Pérez, Candela, and Micol 2009).

Classifying genetic interactions into categories can yield important information concerning the biological pathways in which the genes are taking part. In particular, assumptions can be made concerning whether the interacting genes function in the same pathway or in parallel and convergent pathways. In addition, the pathway order may be defined in the case the interacting genes function in the same pathway, that is if one gene operates upstream or downstream of the other gene. More specifically, considering two single mutants (A and B) and their combined double mutant (AB), an epistatic interaction resembling one mutant (A) could suggest that gene A operates upstream of gene B within a pathway. Similarly, a synergistic interaction resulting to a phenotype stronger or milder than that of the double mutant (AB) would suggest that the two genes (A and B) operate either in successive steps of the same pathway, or in parallel pathways that converge at the same node of a network (Markowetz and Boutros 2015; Pérez-Pérez et al. 2009) (see Fig. 9).

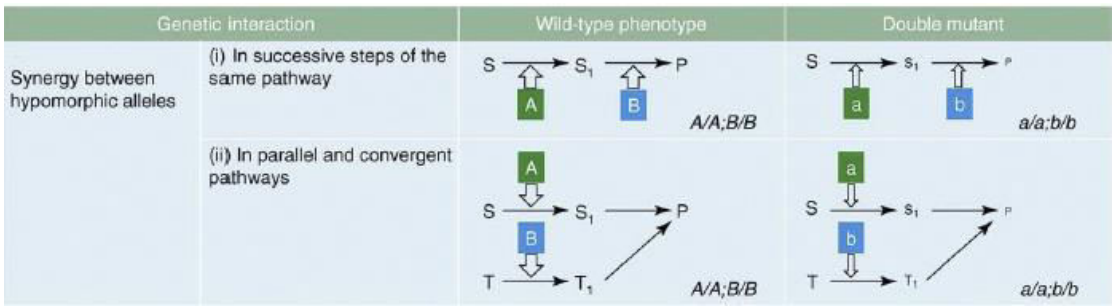


Figure 9. Models for synergistic genetic interactions.

The use of hypomorphic alleles for studying genetic interactions can yield synergistic phenotypes when the two genes (A, B) act in a linear pathway or in two parallel and convergent pathways.

(Reprinted from Trends Genet. 25 (8): 368-76, Pérez-Pérez et al., Copyright 2009, with permission from Elsevier)

1.4.4 The Relationship between Genetic and Physical interactions

The presence of statistical interaction between two genes does not necessarily reflect a direct physical interaction between the two gene products. In fact, only 0.9% of genetic interactions connect proteins that also physically interact, whereas only 10%-20% of interacting protein pairs are encoded by genes that genetically interact (Baryshnikova et al. 2013; Costanzo et al. 2010). This low overlap between genetic and physical interactions is in agreement with the notion that genetic interactions represent a broad spectrum of functional relationships between genes. More often than encoding for physically interacting products, interacting genes may encode proteins that participate in the same regulatory -signaling, transcription factor, metabolic, developmental or epigenetic- network.

1.4.5 SNP-SNP interactions

With a view to shedding light upon the genetic component of complex diseases, the combined effect of SNPs needs to be taken into account, rather than the individual effects of SNPs. For this purpose, interactions occurring between SNPs can be tested on GWAS data, applying more complex statistical analysis, such as logic or logistic regression models.

Interestingly, SNP-SNP interactions have been shown to explain the genetics of some forms of cancer and Crohn's disease (Dinu et al. 2012; Lin et al. 2013; Onay et al. 2006). Moreover, multiple independent SNPs within a single locus were shown to be associated with lipid traits (HDL, LDL, TG, TC) or CAD, doubling the heritability proportion explained for lipid traits (Tada et al. 2014).

In order to discover strongly interacting SNPs, the whole range of potential SNP-SNP interactions need to be tested. Again in this case, the complexity of analysis of higher order interactions limits the search for SNP interactions to pairwise testing for the time being. However, epistasis effects are difficult to detect for complex diseases, even with pairwise testing of interactions between genetic variants. This is because large sample sizes are required in order to gain sufficient statistical power to detect complex and potentially subtle epistatic effects (Lucas et al. 2012; Musameh et al. 2015). Despite the limitations inherent in SNP interaction testing,

studying the co-occurrence of SNPs in patients versus control individuals is considered promising for predicting susceptibility to diseases.

1.4.6 Individualized Medicine: From Personal Genomes to Interactomes

The complete set of interactions between genes and their products is called the "interactome", as an analogy to the genome, proteome and metabolome terminology (Cusick et al. 2005). Following the completion of the human genome sequencing, efforts are currently being made towards the construction of a map of the human interactome network (Human Interactome Project) (Rolland et al. 2014; Rual, Venkatesan, and Hao 2005). Based on the fact that every individual carries a distinct set of genetic variants, which shape its personal genome, transcriptome, proteome and metabolome, it is expected that individual interactomes will be present as well (Zhang, Kuivenhoven, and Groen 2015).

Interactome mapping can be a great benefit for the field of pharmacogenomics, which studies the role of genome in drug response. In cardiovascular disease, currently prescribed medication such as statins has been demonstrated to exert differential effects in patients, depending on the genetic variants they carry (Musunuru 2015). Interactions between these genetic variants are believed to have a substantial impact on the response to medication. Therefore they need to be elucidated, in order to identify the correct dosage for best efficacy, as well as to prevent adverse effects. Thus the current challenge for personalized medicine is the understanding of human interactome, in order to determine the optimal therapy for the disease for each individual patient.

1.5 Mapping Genetic Interactions

Identification of genetic interactions by genetic interaction mapping involves the simultaneous perturbation of two -or more- genes, and assessment of the resulting phenotype in comparison to the one resulting from single perturbation. Genetic interaction mapping methods, which attempt to understand gene function starting with genes of interest and examining the phenotypic consequences of an induced genetic change -mutation, knockout or knockdown- are defined as "reverse genetics". These methods constitute the opposite of "forward genetics" methods,

whereby gene function is studied starting from the phenotypes of interest and subsequently identifying mutants that are responsible for this phenotype (Beltrao, Cagney, and Krogan 2010; Nijman 2011).

Mapping genetic interactions underlying quantitative traits has been performed using different model systems, from unicellular organisms such as yeast (*Saccharomyces*) and bacteria (*Escherichia coli*), to multicellular ones such as worm (*Caenorhabditis elegans*) and fly (*Drosophila*). More recently, genetic interaction studies were carried out using mammalian cells (e.g. mouse and human).

Drosophila has been employed as a model organism to study epistasis since the 1960s by Dobzhansky, Rendel and others. Dobzhansky observed that the combination of certain chromosomes led to synthetic semilethality, that is a slight reduction of viability, which is due to epistatic interactions among genetic variants (Dobzhansky 1965; Spassky, Dobzhansky, and Anderson 1965). The existence of genome-wide deletion libraries in yeast since 2001 (Giaever et al. 2002) made *S. cerevisiae* a convenient model system to study genetic interactions, mainly focusing on synthetic lethality or synthetic sickness (Boone, Bussey, and Andrews 2007). Synthetic lethality and synthetic sickness, in which the joint effect of mutations in two genes is cell death or reduced fitness, respectively, are particularly informative phenotypes, as they can uncover members of the same essential biological pathways (Nijman 2011). These fitness readouts were used in *S. cerevisiae* to perform a global mapping of its interaction network by applying synthetic genetic array (SGA) analysis, an approach to systematically generate and isolate double mutant strains (Tong 2004).

Whereas in yeast comprehensive deletion libraries have made it possible to screen for genetic interactions in a systematic, high-throughput way, by completely discarding the product of the genes under study ("knockout"), such approaches are not feasible for more complex organisms. The reason for this is mainly a scaling issue; whilst in *S. cerevisiae* there are approximately 6,000 genes, *D. melanogaster* genome contains around 17,000 genes and both *C. elegans* and human have around 21,000 genes, which would lead to a huge number of combinations. Therefore, an approach that takes advantage of RNA interference (RNAi) has been used in

multicellular organisms, examining the effect of the reduction of gene products (knockdowns rather than knockouts).

1.5.1 RNA INTERFERENCE

The mechanism of RNAi is based on the enzymatic cleavage of long double-stranded RNAs into short 21-nucleotide long RNAs, named “short interfering RNAs” (siRNAs). Following the “dicing”, the siRNA triggers the destruction of the homologous mRNA, which results in the so-called “silencing” of the gene, namely the inhibition of its expression. Thus, the function of a gene can be studied by observing the phenotype resulting from its RNAi-induced silencing.

In contrast to gene knockouts, as in the case of deletion libraries in yeast, whereby a gene product is completely abolished, RNAi-mediated gene knockdowns frequently reduce protein levels only up to a certain extent. The incomplete knockdown of a gene, which is termed DAmP (Decreased Abundance by mRNA Perturbation), generates a partial phenotype (hypomorph) that would not be observed upon gene knockout. Importantly, partial gene product depletion generated by RNAi can be informative for genetic interactions. This is suggested by studies performed in yeast, in which the use of hypomorphic (DAmP) alleles revealed biologically meaningful genetic interactions (Roguev et al. 2013; Schuldiner et al. 2005). Moreover, phenotypes generated by RNAi-induced incomplete gene knockdowns very closely resemble those in patients carrying partial loss-of-function (hypomorphic) mutations.

The construction of siRNA libraries for model organisms such as *Drosophila* and *C. elegans*, as well as for mammalian cells, in combination with a recent advance in high-throughput screening technologies has enabled the appliance of RNAi for genetic interaction mapping.

1.5.2 HIGH-THROUGHPUT SCREENING USING RNAI

Importantly, a microscopy platform was established at the European Molecular Biology Laboratory (EMBL) for high-throughput RNAi screens using fluorescence-based imaging assays in cells (Pepperkok and Ellenberg 2006). This platform was based on the development of an RNAi-based reverse transfection protocol on pre-

spotted microarrays of 384 samples, using mammalian tissue culture cells (Erfle et al. 2007, 2008). The aforementioned platform has already been used in genome-wide RNAi screens studying cell division (Neumann et al. 2011), signal transduction (Ritzerfeld et al. 2011) and secretion (Simpson et al. 2012). In addition, the above platform has been used in targeted RNAi screens studying cholesterol regulation (Bartz et al. 2009; Blattmann et al. 2013).

1.5.3 High-Content Genetic Interaction Screening Using co-RNAi

To identify genetic interactions using RNAi, multiple genes are targeted by siRNAs simultaneously, in an approach described as "combinatorial RNAi" (co-RNAi). Combinatorial RNAi was first used to study genetic interactions in *C. elegans* (Lehner, Tischler, and Fraser 2006; Tischler et al. 2006) and *Drosophila* (Billmann et al. 2016; Horn et al. 2011), while more recently it was employed for genetic interaction mapping in human cells (Laufer et al. 2013). Over the last decade, the technology of combinatorial RNAi has been employed for mapping genetic interactions in mammalian cells in a systematic, automated, large-scale manner (Barr and Bakal 2015; Bassik et al. 2013; Kampmann, Bassik, and Weissman 2013; Laufer et al. 2013; Roguev et al. 2013; Wang et al. 2014). High-throughput gene-gene interaction screens are performed using 96- or 384-well plates and microarrays, robotics, as well as automated imaging and image analysis.

For the time being, gene-gene interaction testing has been limited to pairwise, due to the complexity of statistical analysis required for higher-order interactions, although interactions among a number of genes may account for an observed phenotype. Specifically, pairwise RNAi-mediated perturbations of gene function ("double knockdowns") have been employed in studies of genetic interactions (GIs) (Barr and Bakal 2015; Bassik et al. 2013; Billmann et al. 2016; Horn et al. 2011; Laufer et al. 2013; Roguev et al. 2013; Wang et al. 2014). Such reports have demonstrated the potential of this approach to reveal novel genetic interactions, discover protein complexes, functional modules and pathways, and to infer gene function.

1.5.4 MAPPING GIS IN CHOLESTEROL HOMEOSTASIS WITH CO-RNAI

In our laboratory, an RNAi-based functional profiling of GWAS-derived loci, associated with lipid traits, CAD and/or MI was previously performed. This work was aimed towards gaining a mechanistic understanding of how the loci identified affect cholesterol regulation (Blattmann et al. 2013). The study resulted in the identification of 55 genes that had an effect on LDL internalization and/ or cellular cholesterol levels.

However, most of the screen "hits" did not have a strong effect on cholesterol regulation. This suggests that the combinatorial -rather than the individual- function of genes might regulate cholesterol homeostasis. The complexity of cholesterol regulation is in agreement with the complexity of CVD, for which dysregulated cholesterol is a major risk factor. Keeping all the above in mind, we decided to perform a co-RNAi-based interaction screen, in order to explore possible interactions between the genes identified in the aforementioned study.

2 AIMS OF THE STUDY

The pathogenesis of coronary artery disease (CAD), the most common form of cardiovascular disease (CVD), results from elevated levels of LDL-cholesterol in the bloodstream, a consequence of defective cholesterol-regulatory mechanism in the cells. CAD is a complex disease, believed to result from multiple combinations of gene-gene interactions. Therefore, multiple inherited DNA variants are assumed to contribute to CAD risk, each of which has a small individual effect on disease risk. The study of Blattmann et al. identified a number of disease-regulated genes with an effect on cholesterol regulation, most of which however did not have a strong individual effect. The question that arises therefrom is if and how these cholesterol-regulating genes interact with each other, in order to collectively control cholesterol homeostasis. Furthermore, it is interesting to discover how these genetic interactions may contribute to disease risk.

To address these questions, the aims of the present study were the following:

i. to screen for and identify gene-gene interactions among cholesterol regulators
ii. to reveal how these gene-gene interactions contribute to cellular cholesterol homeostasis and
iii. to define how these gene-gene interactions are linked to CVD.

3 RESULTS

The work described here is divided into three parts. In the first part (3.1), the combinatorial RNA interference (co-RNAi) approach is described that was used to identify gene-gene interactions with an effect on LDL endocytosis. After establishing the high-throughput microscopy-based approach (3.1.1) and determining the optimal siRNA concentrations (3.1.2), candidate genes were selected (3.1.3), as well as the corresponding siRNAs (3.1.4) for the screen. The results of the co-RNAi screen are described (3.1.5) and the gene interactions identified are shown (3.1.6). With a number of secondary experiments (3.1.7) the identified gene interactions were validated (3.1.8). The validated gene interactions were clustered and finally used to construct a model network (3.1.9). Some of these interactions were followed up (3.1.10). In the second part, the effect of mutations of one prominent screen interactor on LDL uptake was studied (3.2). In the third and last part, the co-occurrence of SNPs of the candidate genes tested in the co-RNAi screen was examined in a cohort of >4,000 individuals (3.3).

3.1 COMBINATORIAL RNAI SCREEN

3.1.1 ESTABLISHMENT OF A HIGH-THROUGHPUT MICROSCOPY-BASED APPROACH TO SYSTEMATICALLY IDENTIFY GENE INTERACTIONS AFFECTING CHOLESTEROL HOMEOSTASIS

With a view to defining gene networks underlying cellular cholesterol homeostasis, I performed pairwise double knockdowns *in vitro* between 30 disease-associated genes, by reverse transfecting HeLa Kyoto cells on siRNA-spotted arrays (LabTeks), as described (see 7.1.2.2.1) (Erfle et al. 2007, 2008). For the analysis of knockdown effects, I used a cell-based assay, previously established in the lab to measure the effect of the gene knockdowns on cholesterol regulation (Bartz et al. 2009). This is a quantitative LDL-uptake assay, which measures the ability of tissue culture cells to endocytose fluorescently (DiI) labeled LDL (see 7.1.5.1). Uptake of fluorescently labeled LDL from the siRNA-transfected cells was determined using high-content automated microscopy (see 7.4.1.2.1) (Bartz et al. 2009; Erfle et al. 2007). The experimental workflow is illustrated in Figure 10.

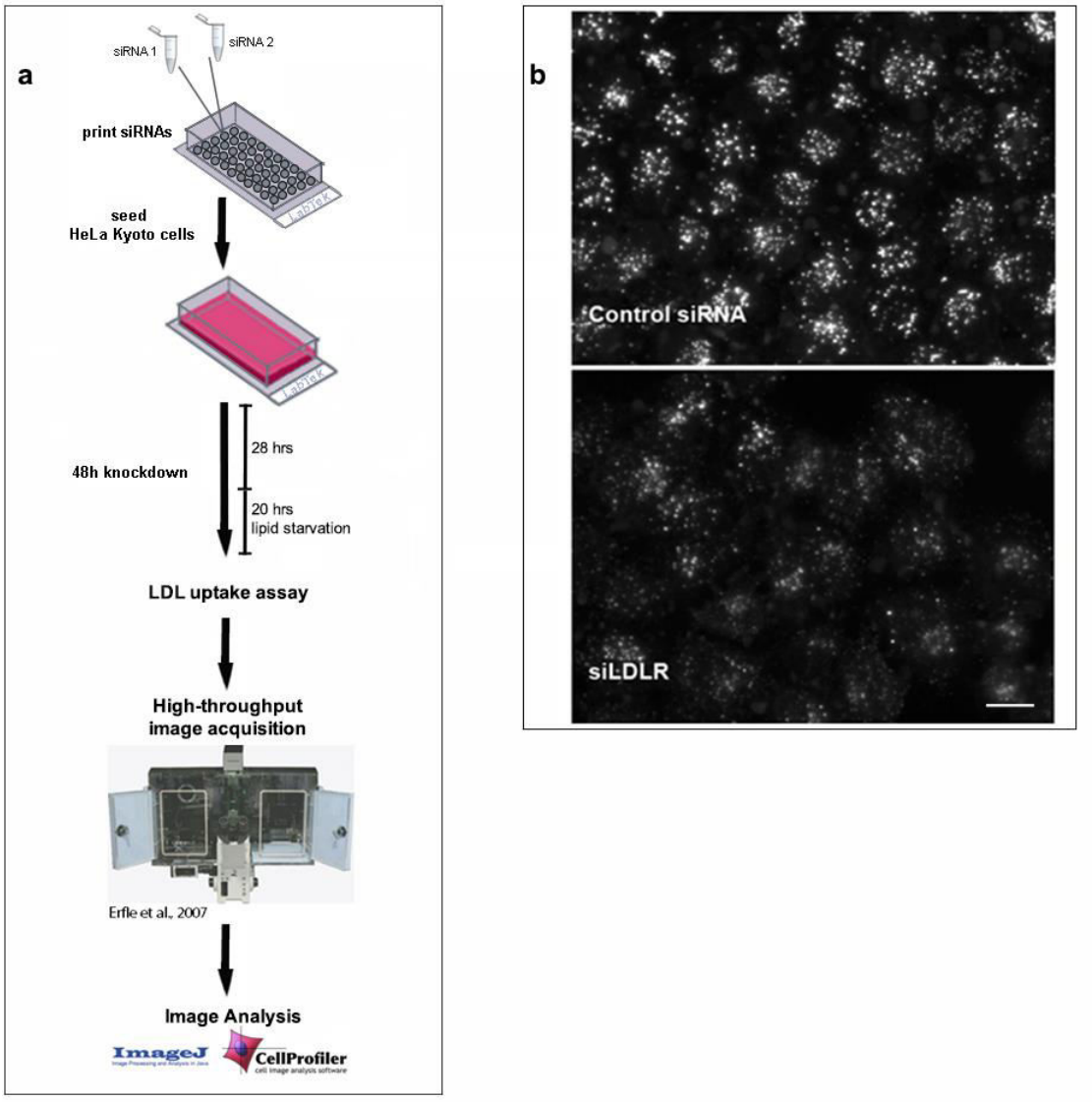


Figure 10. The experimental workflow followed in the high-throughput screening for cholesterol interactors

a. For the co-RNAi interaction screen, cell microarrays (LabTeks) were printed with 2 siRNAs per spot, and afterwards HeLa Kyoto cells were seeded on the spotted LabTeks. After a 48-hour knockdown, the LDL-uptake assay was performed in the LabTeks, and images of the transfected cells were acquired in a widefield microscope in a high-throughput manner. Finally, the images were analysed using image analysis softwares (FiJi-ImageJ and CellProfiler). b. Shown are representative images of LDL-uptake assay: cells transfected with negative control siRNA show a normal uptake of fluorescently labeled LDL (top image), whereas cells transfected with siRNA targeting the LDL receptor show a significant decrease in fluorescence signal of endocytosed LDL (bottom image) (scalebar=10 µm).

3.1.2 DETERMINING THE OPTIMAL SIRNA CONCENTRATION FOR THE DOUBLE KNOCKDOWNS

In order to determine the optimal siRNA concentration for the double knockdowns on LabTeks, two conditions were examined in a pilot experiment. For this purpose, single knockdowns as well as double knockdown of LDLR and DAB2 were performed on spotted LabTeks, using either 30pmol or 15pmol total siRNA (per well of a 384-low-volume-plate used for the spotting of the LabTek, see (Erfle et al. 2007)). NEG9 was used as negative control siRNA and INCENP as transfection control siRNA (see also section 7.1.2.2.1.1). The selection of the two target genes, LDLR and DAB2, was based on the inclusion of one gene with strong knockdown effect and one with mild knockdown effect. LDLR is the key regulator of LDL internalization, therefore its knockdown leads to drastic impairment of LDL endocytosis, whereas DAB2 has an auxiliary role in LDL internalization (Chetrit, Ziv, and Ehrlich 2009; Maurer and Cooper 2006; Mishra et al. 2002), thus its individual knockdown has no significant effect on LDL internalization (Eden et al. 2007; Wei, Hemani, and Haley 2014). The transfection control siRNA targets *INCENP*; Inner Centromere Protein, which is necessary for chromosome segregation during mitosis, and therefore its knockdown results in multi-nucleated cells and it has been previously used in siRNA screens (Erfle et al. 2007).

The knockdown effect on LDL internalization was evaluated using the LDL-uptake assay (for details see 7.1.6.1). The transfection efficiency in the two conditions was also evaluated by quantifying the number of cells showing the characteristic INCENP phenotype, upon knockdown with INCENP.

Although there was not significant difference between the two conditions on LDL uptake (see Fig. 11), the INCENP phenotype (see Fig. 12) was more penetrant with the higher siRNA concentration: 16% of the cells showed the phenotype with 15pmol siRNA versus 59% with 30pmol. Therefore, the higher siRNA concentration (30 pmol) was selected for the co-RNAi screen, in order to achieve robust knockdowns.

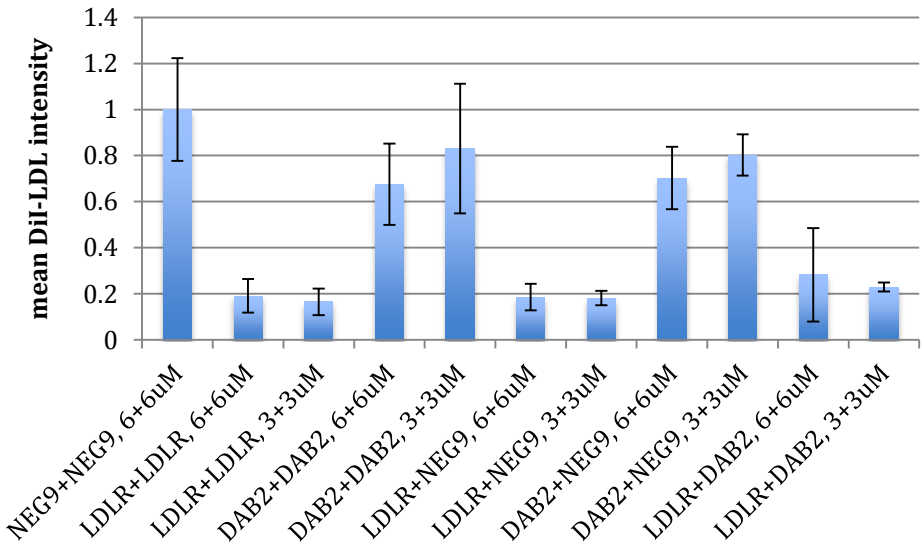


Figure 11. Results of the experiment for siRNA concentration optimization for double knockdowns
Shown in this graph are the effects of knockdowns on LDL uptake, using either 15pmol (3+3uM) or 30 pmol (6+6uM) total siRNA. The mean fluorescent intensities of endocytosed DiI-LDL are normalized to the negative control siRNA; NEG9. Error bars represent the standard deviation (n=8).

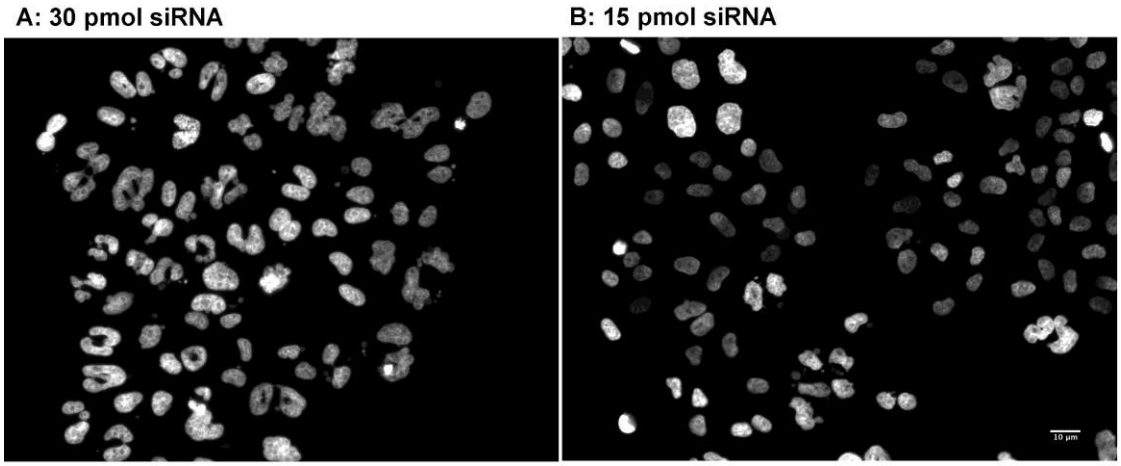


Figure 12. Cells showing the characteristic INCENP phenotype
Shown are cells transfected with either 30 pmol (A) or 15 pmol (B) siRNA targeting the Inner Centromere Protein (*INCENP*) gene. Knockdowns of this gene results in mitotic arrest, strange nuclear shape, binuclear cells or polylobed nuclei, due to spindle assembly defects (Neumann et al. 2011).

3.1.3 CANDIDATE GENE SELECTION STRATEGY AND RATIONALE

Thirty candidate genes were selected for combinatorial RNAi screening (Figure 3.2); all of them were identified by GWAS as associated with blood lipid levels, coronary artery disease (CAD), and/or myocardial infarction (MI) (Tbl. S3). Of these, 28 genes were previously shown to affect low-density lipoprotein (LDL) internalization and/or cellular levels of free cholesterol in Blattmann et al. (Blattmann et al. 2013). It should be noted that the genes tested in the aforementioned study originate from GWA studies, which are only powered to detect common variants-SNPs with a minor allele frequency (MAF)>5%. In order to more efficiently address the missing heritability problem that was described previously (see also section 1.1.3.2.1), genes with rare variants associated to lipid traits were also included in the screen. Besides, the inclusion of the total of 55 "hits" of the Blattmann et al. study in the combinatorial RNAi screen would end up in 3025 (55x55) combinations, without taking into account the controls. Therefore, a filtering of these hits had to be done. For these reasons, the results of the Blattmann et al. study were filtered by comparing them with those of an Exome Chip for lipid traits (Peloso et al. 2014) (see Tbls. S1 and S2). This approach narrowed down the candidate gene number to 30 and the pairwise combinations to 900. A relatively small gene combination set was preferable in order to perform the experiments in many replicates, which led to more reliable data.

The Blattmann et al. study utilized RNAi in an effort to complement GWAS data with functional information. Until 2009 when the aforementioned study started, 22 GWAS had been published analyzing association with serum lipid traits or cardiovascular disease. For the Blattmann et al. study, 56 from the 64 identified loci were selected, and 133 genes that represented these 56 loci were analyzed for a role in cholesterol regulation. This study led to the identification of 55 genes with a functional effect on LDL uptake and/ or cellular free cholesterol levels.

In the second study that was utilized for the candidate gene selection, a total of approximately 70,000 individuals were genotyped for LDL, HDL, triglycerides (TG) and total cholesterol (TC) traits. The genotyping was performed with an Illumina HumanExome genotyping array ("the Exome Array" or "Exome Chip") built with variants discovered upon exome-sequencing of ~12,000 individuals (Tbl. S1, Single-variant-based analyses performed by Anirruddh Patel and other, Kathiresan lab,

Massachusetts General Hospital, Boston). This study revealed 165 “hits”-genes with variants significantly associated with lipid traits. A large part of this study has been subsequently published (Peloso et al. 2014).

The “hits” from the two aforementioned studies were juxtaposed (Tbl. S2), and the 21 overlapping genes were selected (*APOB*, *APOE*, *BAZ1B*, *BCAM*, *BCL7B*, *CBLC*, *CELSR2*, *HAVCR1*, *HMGCR*, *LDLR*, *LPL*, *MLXIPL*, *MYBPHL*, *NCAN*, *PCSK9*, *PVRL2*, *SIK3*, *TM6SF2*, *TMEM57*, *TOMM40* and *ZNF259*). Additionally, to the candidate gene set a few genes were added that did not overlap between the two studies, as they were considered significant. First, seven genes from the hits of Blattmann et al., that did not overlap with the Exome Chip hits were selected, for the reasons listed below. Five of these genes -*CXCL12*, *SORT1*, *FAM174A*, *WDR12* and *SEZ6L*- were the strongest hits in Blattmann et al. (Blattmann et al. 2013). *LDLRAP1/ARH* gene was included as it has a well-described role in LDLR endocytosis, and its mutations are responsible for autosomal recessive hypercholesterolemia.

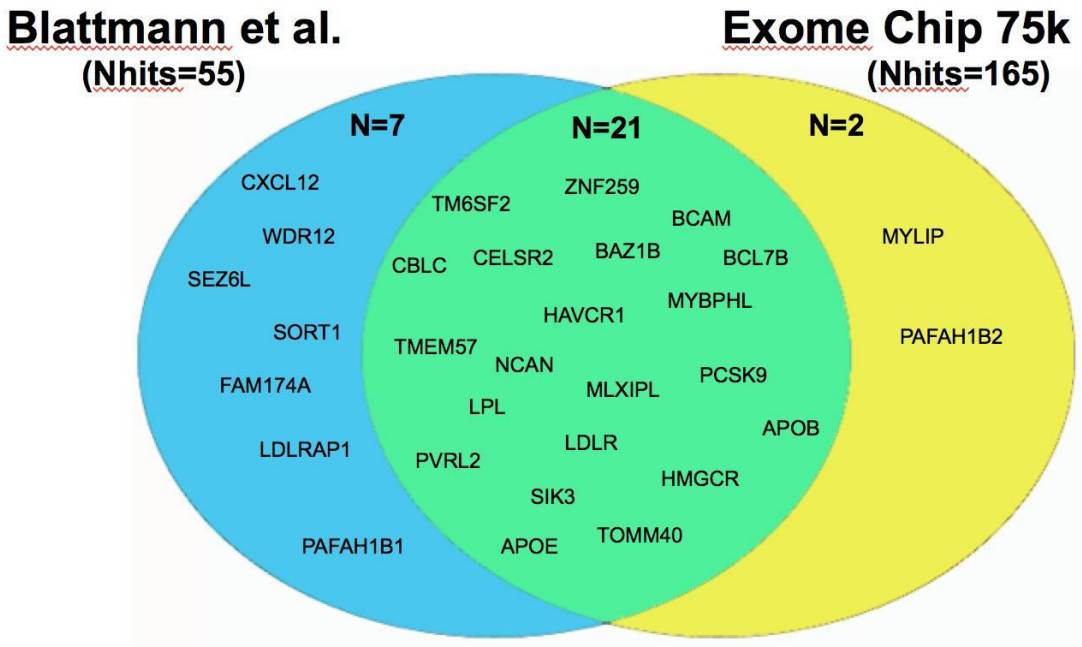


Figure 13. Final selection of 30 candidate genes for the interaction screen.
In total, 30 genes were selected to be tested for gene-gene interactions with co-RNAi. Of those 30 genes, 21 overlapped between the Blattmann et al. study and the Exome chip (green), seven were additionally selected from the Blattmann et al. (blue), and two were selected from the Exome Chip (yellow).

PAFAH1B1 was selected as it may have a significant role in cholesterol regulation, based on a number of unpublished experiments by Blattmann et al. Secondly, two genes that were hits in the Exome Chip but were not analyzed in Blattmann et al. were selected; of these, *MYLIP*/*IDOL* has a well-described role in endocytosis and degradation of LDLR (Zelcer 2009), and *PAFAH1B2* encodes a subunit of the PAFAH1B complex that is formed together with PAFAH1B1, which was shown to affect LDL internalization and/or cellular levels of free cholesterol (Blattmann et al. 2013). So, the final selection consisted of 30 candidate genes (see Fig. 13 and Tbl. S4). The reported function of the candidate genes, as well as their link -if existing- to cholesterol regulation or lipid homeostasis are shown in Tbl. S5.

3.1.4 siRNA SELECTION

The combinatorial RNAi screen contained 30 siRNAs each targeting the one of the selected genes, as well as two negative control siRNAs, one positive control siRNA for Filipin assay and one transfection control siRNA.

In my screen, gene knockdowns were performed using only one Silencer Select siRNA per gene. For the 28 genes that were previously analyzed in Blattmann et al. (Blattmann et al. 2013), siRNAs were selected from this study. In Blattmann et al., three-five independent siRNAs were tested per gene. For my screen, I chose the siRNA that had the strongest effect on either of the two assays performed (LDL-uptake assay preferred over Filipin assay). For the two genes that were not analyzed in Blattmann et al. (*MYLIP*, *PAFAH1B1*), the Silencer Select siRNAs that targeted most gene transcripts were selected. The siRNAs used and the selection criteria are shown in detail in Tbl. S6.

All siRNA sequences were mapped to the human reference genome GRCh37 (Ensembl 75) using the in-house software tool Bluegecko (J.K. Hériché, unpublished). With this tool the number of targeted protein-coding transcripts was evaluated and unspecific siRNAs that target other human mRNAs were identified. Moreover, the siRNAs were evaluated for mismatches to the reference sequence of the respective target gene or for targeting transcripts not anymore considered as protein coding.

3.1.5 RESULTS OF THE COMBINATORIAL RNAI SCREEN

Using the selected siRNAs, the double knockdowns were performed on HeLa Kyoto cells for all the combinations between the 30 candidate genes on spotted LabTeks, as described (for details see 7.1.2.2.1.2). The LDL-uptake assay was performed to measure the effect of the combined knockdowns on LDL internalization and the fixed cells were imaged using a widefield Olympus Scan^R microscope (see also 7.1.5.1 and 7.1.4.2). Internalized DiI-LDL intensity was measured in at least 6 biological replicas (after excluding those which did not pass quality control) per spotted array, acquiring one image per spot.

All images from the screen were analyzed performing automated image analysis with a Cellprofiler pipeline, in which DiI-LDL fluorescence signal was quantified within intracellular areas-"dots" (for details see 7.4.1.2.1). From each cell, the following parameters were measured:

(i) total intensity above local background within all dots in one cell ("integrated intensity"),

(ii) mean intensity above local background within these dots per cell ("mean intensity"),

(iii) number of dots per cell and

(iv) mean area covered by dots per cell.

The total intensity corresponded to the total uptake of LDL, while the mean intensity was assumed to correspond to LDL concentration within the segmented structures.

3.1.5.1 PRIMARY ANALYSIS; CONTROL-BASED NORMALIZATION

The readout chosen for quantifying the knockdown effect and afterwards for defining the interactions was the mean total fluorescence intensity of internalized DiI-LDL per cell, per image. A robust Z score was calculated by normalizing the total intensity of a treatment to the controls of the identical plate;

$$robustZscore = \frac{treated - median(Ctrls)}{mad(Ctrls)}$$

(Birmingham et al. 2009; Malo et al. 2006). Thus, a positive robust Z-score meant enhancement of LDL uptake, whereas negative meant inhibition. The robust Z-score was calculated using a self-written R-script (for details see 7.4.2.1). Within this R-

script, a median robust z score per treatment was calculated for each treatment, by taking the median of robust z scores of same treatments across different biological replicates.

The analysis of the screen resulted to both predicted, as well as unpredicted outcomes. For instance, LDLR knockdown had a negative effect on LDL uptake in combination with most other genes, which was expected considering that in absence of LDL receptors on the plasma membrane LDL uptake is significantly hindered. Interestingly, however, LDLR double knockdown with HMGCR or PVRL2 did not have a strong effect on LDL uptake. In these two cases, the effect of LDLR knockdown on LDL uptake seemed to be counteracted by the co-knockdown of the second gene. A heatmap representation of the screen results can be seen in Figure 14.

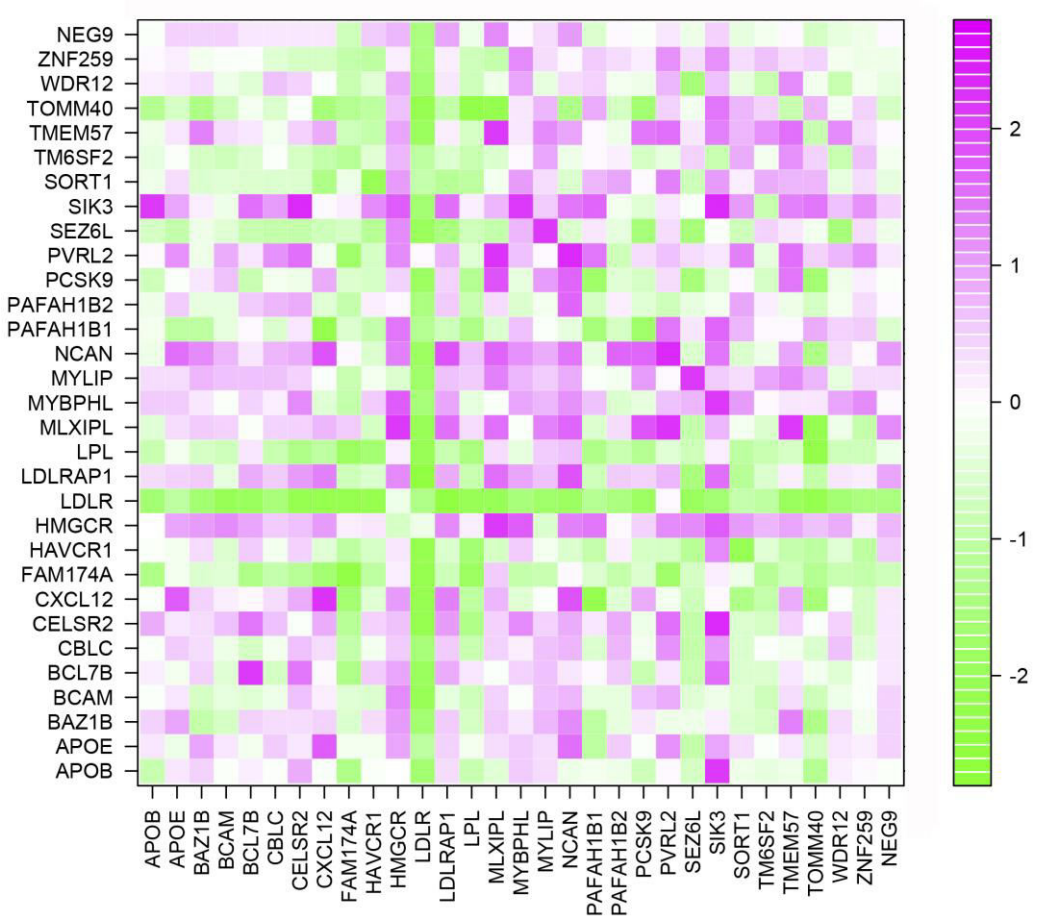


Figure 14. Heatmap matrix representation of the combined effects of double knockdowns on DiI-LDL uptake. The colours represent effect size on the linear scale: positive effects are in magenta and negative in green. The readout is the median robust Z score of internalized DiI-LDL intensity per treatment; robust Z score=(treated-median(controls))/mad(controls). Numeric data are shown in Tbl. S19.

3.1.5.2 ALTERNATIVE ANALYSIS; SAMPLE-BASED NORMALIZATION

In an alternative analysis the robust Z score was calculated by normalizing the total intensity of a treatment to all the treatments of the identical plate (Birmingham et al. 2009; Malo et al. 2006). The transfection controls (Incenp) and the positive controls (Npc1, Ldlr) were excluded from the normalization.

$$robustZscore = \frac{treated - median(all)}{mad(all)}.$$

The robust Z score was calculated using a self-written R-script (for details see 7.4.2.2). Within this R-script, a median robust z score per treatment was calculated for each treatment, by taking the median of robust z scores of same treatments across different biological replicates.

This analysis approach is based on the assumption that most of the treatments do not display biological effect in the analyzed assay. A heatmap representation of the screen results can be seen in Figure 15.

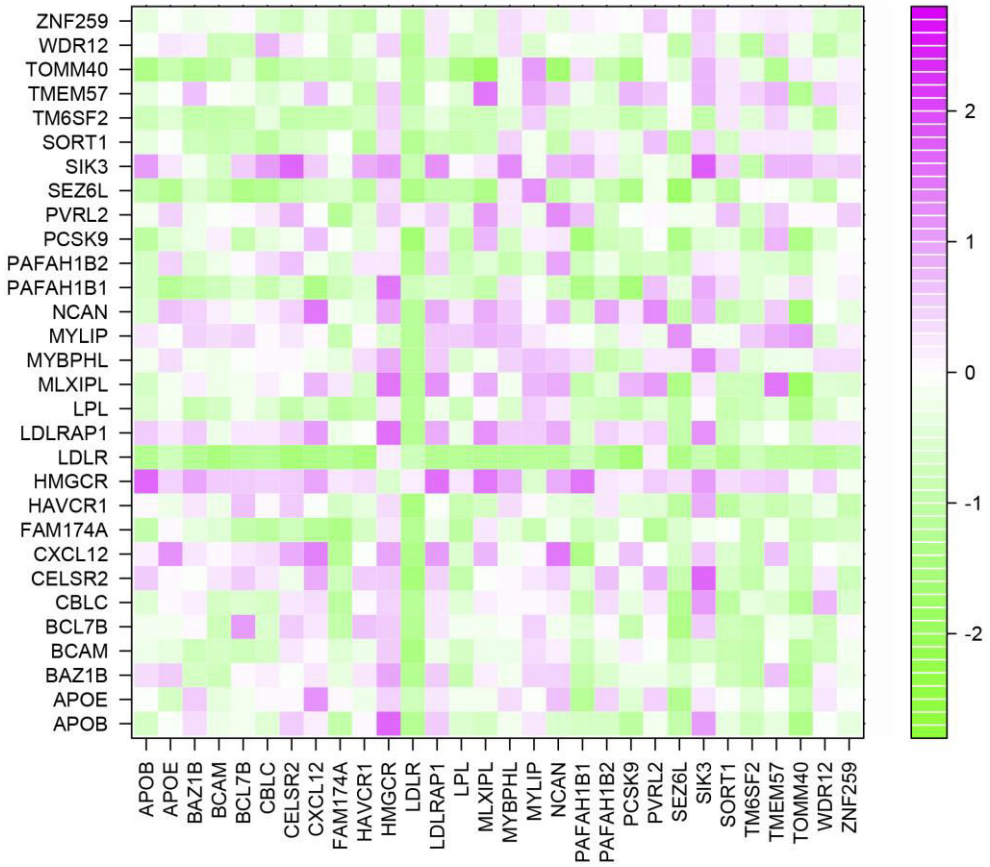


Figure 15. Heatmap matrix representation of the combined effects of double knockdowns on DiI-LDL uptake, calculated with the alternative approach. The colours represent effect size on the linear scale: positive effects are in magenta and negative in green. The readout is the median robust Z score of internalized DiI-LDL intensity per treatment; robust Z score=(treated-median(all))/mad(all).

3.1.6 Defining Gene-Gene Interactions among the results of the screen

3.1.6.1 primary approach; control-based analysis

To test for gene-gene interactions among the results of the combinatorial RNAi screen, the effect on LDL uptake of the combined transfection with two siRNAs targeting two different genes A and B was compared to the additive effect of the two single transfections. In single transfections the respective siRNAs were co-transfected together with the negative control siRNA. For this, a multiple linear regression model was applied to analyze the screen data using an R-script that was written with the help of Bernd Klaus from the Huber Group at EMBL (for details see 7.4.3). Thereby, the additive effect of two single knockdowns A and B {($Gene_A$+control)+($Gene_B$+control)} was compared to the respective double knockdown effect ($Gene_A$+$Gene_B$). Thus, a double knockdown phenotype was considered indicative of a genetic interaction when it deviated from the additive effect of the individual knockdowns.

This analysis was based on Fisher's definition of gene interaction -termed "epistasis"; "Epistasis is a deviation from the additivity in the effect of alleles at different genetic loci with respect to their contribution to a quantitative phenotype" (Fisher 1918) (see also section 1.4.2). In the regression model used, the interaction effect of two genes was calculated from the difference between the additive effect of the two single knockdowns and that of the double knockdown (Axelsson et al. 2011; Horn et al. 2011). Finally, a p-value was calculated from the t-value of the regression model and the p-value was adjusted for multiple comparisons using the false discovery rate (fdr) method (Benjamini and Hochberg 1995). The significance threshold for gene interactions was set at $pValue_{fdr}<10^{-2}$.

With this analysis, 35 gene interactions were identified. The identified interactions and their interaction effect and significance (p-value) are shown in Tbl. 4 and Fig. 16.

	Gene1	Gene2	robustZscore	Interaction Value	pVal (fdr)
1	PAFAH1B1	TOMM40	0.89	2.29	6.54E-05
2	APOB	HMGCR	2.80	2.18	4.05E-03
3	SORT1	TOMM40	0.77	1.84	3.90E-03
4	PAFAH1B1	SIK3	1.62	1.79	2.52E-03
5	PCSK9	TMEM57	1.46	1.76	2.37E-03
6	MYLIP	SEZ6L	2.18	1.75	2.22E-03
7	HMGCR	PAFAH1B1	1.45	1.74	6.59E-03
8	CELSR2	SIK3	2.30	1.60	8.09E-03
9	MYBPHL	SIK3	2.17	1.59	6.59E-03
10	PVRL2	SORT1	1.36	1.54	5.88E-03
11	LDLR	PVRL2	0.08	1.48	4.23E-03
12	LDLR	NCAN	-1.86	-1.28	7.64E-03
13	APOE	MLXIPL	0.35	-1.40	7.25E-03
14	HAVCR1	MYLIP	-0.13	-1.40	5.60E-03
15	CELSR2	LPL	-1.40	-1.43	7.53E-03
16	LDLR	MLXIPL	-2.03	-1.49	2.97E-03
17	BCAM	HAVCR1	-0.51	-1.56	2.97E-03
18	NCAN	TOMM40	-1.40	-1.56	7.17E-03
19	NCAN	SEZ6L	-0.59	-1.60	5.54E-03
20	HAVCR1	ZNF259	-1.01	-1.62	3.23E-03
21	LDLRAP1	SORT1	-1.11	-1.65	2.57E-03
22	LDLRAP1	LPL	-0.81	-1.68	3.69E-04
23	HAVCR1	LPL	-1.65	-1.75	4.01E-04
24	CXCL12	PAFAH1B1	-2.17	-1.78	3.69E-04
25	HAVCR1	TMEM57	-0.96	-1.82	4.62E-04
26	BCAM	LDLRAP1	-0.40	-1.83	3.58E-04
27	LDLR	LDLRAP1	-2.44	-1.86	9.34E-05
28	HAVCR1	SEZ6L	-1.19	-1.91	3.39E-04
29	HAVCR1	LDLRAP1	-0.56	-2.16	1.03E-05
30	HAVCR1	NCAN	-0.51	-2.19	2.18E-05
31	LDLRAP1	SEZ6L	-1.46	-2.20	1.60E-05
32	HAVCR1	SORT1	-2.09	-2.24	9.34E-05
33	HAVCR1	MLXIPL	-0.63	-2.31	1.03E-05
34	MLXIPL	SEZ6L	-0.91	-2.40	2.18E-05
35	MLXIPL	TOMM40	-2.14	-2.67	1.03E-05

Table 4. Hits of the combinatorial RNAi screen

The 35 gene pairs that were identified as interacting in the combinatorial RNAi screen. The statistical significance threshold was set at p-value (fdr)<10^-2. Shown are the median robustZscores, the interaction effects calculated with the linear regression model, as well as the p-values after fdr correction for each gene pair.

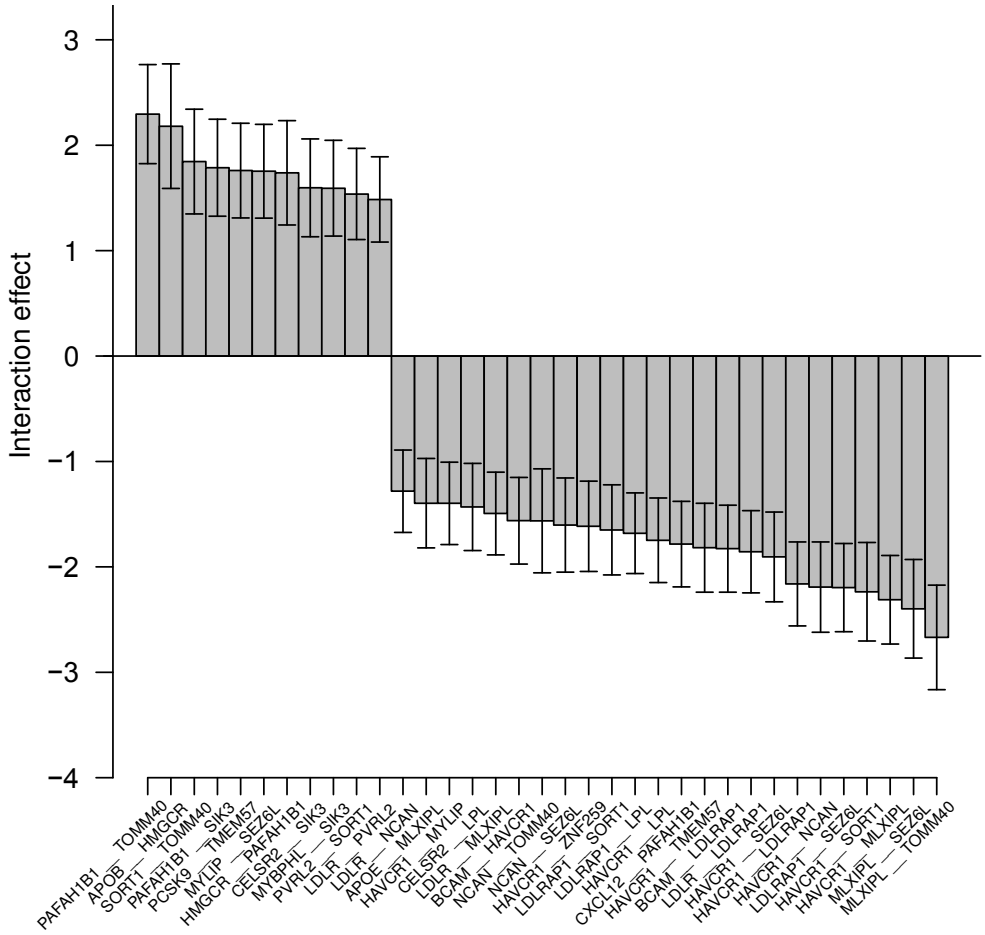


Figure 16. Hits of the combinatorial RNAi screen

The 35 gene pairs that were identified as interacting from the combinatorial RNAi screen. The statistical significance threshold was set at p-value (fdr)<10^-2. The Y axis shows the estimated interaction effect of the robust linear model.

3.1.6.2 ALTERNATIVE APPROACH; SAMPLE-BASED ANALYSIS

In an alternative statistical analysis to test for gene-gene interactions among the results of the combinatorial RNAi screen, the double knockdown effect ($Gene_A$+$Gene_B$) was compared to the additive effect of two single knockdowns A and B {($Gene_A$+any)+($Gene_B$+any)}, in which case as single knockdown I considered the combination of the respective gene with any other gene.

This analysis was based on the observation that the knockdown effect of an siRNA was sometimes milder when it was co-transfected with a second siRNA, as compared to when transfected alone or with negative control siRNA. Therefore, it was considered as more "fair" to compare double knockdowns ($Gene_A$+$Gene_B$) with double knockdowns ($Gene_A$+any) or ($Gene_B$+any), instead of comparing with "single" knockdowns ($Gene_A$+control) or ($Gene_B$+control), assuming that the control siRNA has no effect.

With this analysis, seven gene-gene interactions were identified at $pValue_{fdr}<5x10^{-2}$; (*MYLIP*+*SEZ6L*), (*PAFAH1B1*+*TOMM40*), (*SORT1*+*TOMM40*), (*LDLR*+*PVRL2*), (*APOE*+*CXCL12*), (*CXCL12*+*PAFAH1B1*), (*MLXIPL*+*TOMM40*), and only two at $pValue_{fdr}<10^{-2}$; (*MYLIP*+*SEZ6L*), (*PAFAH1B1*+*TOMM40*).

The identified interactions and the interaction effect and significance of the interaction (p-value) are shown in Figure 17.

	Gene1	Gene2	Interaction Value	pVal (fdr)
1	APOE	CXCL12	1.01	3.90E-02
2	CXCL12	PAFAH1B1	-1.13	1.84E-02
3	LDLR	PVRL2	1.10	2.27E-02
4	MLXIPL	TOMM40	-1.25	2.80E-02
5	MYLIP	SEZ6L	1.83	8.09E-05
6	PAFAH1B1	TOMM40	1.40	6.63E-03
7	SORT1	TOMM40	1.36	2.27E-02

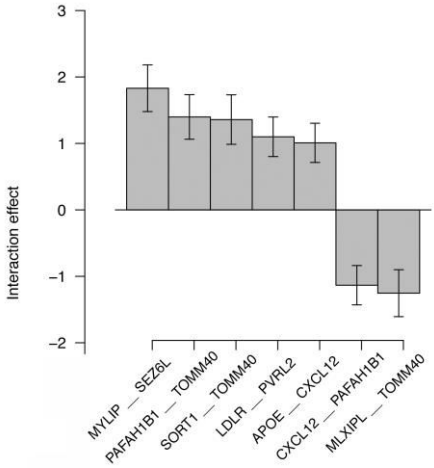


Figure 17. The seven gene pairs were identified as interacting in the combinatorial RNAi screen (alternative analysis).

The statistical significance threshold was set at p-value (fdr)<10^{-2}. Shown in the table are the interaction effects calculated with the linear regression model, as well as the p-values after fdr correction for each gene interaction. The Y axis of the figure shows the estimated interaction effect of the robust linear model.

3.1.6.3 OVERLAP OF HITS BETWEEN TWO ANALYSES

At pValue(fdr)< 10^{-2} , the hits that were identified in the two analysis overlapped 100%. At pValue(fdr)< $5x10^{-2}$, seven hits were identified with the second analysis, that also overlapped 100% with the (99) hits that were identified at the same threshold with the first analysis (not shown here).

3.1.7 SECONDARY EXPERIMENTS

3.1.7.1 REPLICATION AND VALIDATION OF SCREEN RESULTS

In order to perform a number of secondary experiments that could not be done with the microarrays, the 35 double knockdowns that emerged as screen hits (primary analysis) were repeated in a larger format using liquid-phase forward transfection (for details see section 7.1.2.1). The purpose of doing this was to measure the knockdown efficiency of the siRNAs used, as well as to determine if these double knockdowns had an effect on LDLR mRNA and/or protein levels, and most importantly to validate the hits with a different method. To this end, the single knockdowns (GeneA+control) for each interacting gene as well as the double knockdowns (GeneA+GeneB) were performed, and the same analysis was conducted as for the screen (for details see sections 3.1.4 and 3.1.5).

Liquid phase transfections resulted in more penetrant and stronger phenotypes. Thus adjustment of the siRNA concentrations was necessary to mimic the experimental setup of the screen (see Fig. S1).

Of the 32 screen hits, 20 were validated with liquid-phase transfection. These 20 hits had a statistically significant interaction effect, which was similarly up- or down-regulating as in the screen. One more interaction (*LDLR*+*HAVCR1*) that had not reached statistical significance in the screen, but seemed interesting to us (see section 3.3) was independently confirmed. The 21 interactions that were confirmed with liquid-phase transfection are shown in Tbl. 5 and Fig. 18 (see also Fig. S2). The siRNA amounts and knockdown efficiencies for the single and double knockdowns are shown in Tbls. S7 and S8, respectively, and the siRNA amounts used for double knockdowns in the validation experiments are shown in Tbl. S9. The results of the validation experiments are shown on Tbl. S20.

	Treatment	Robust Z score	Interaction value	pVal (fdr)
1	LDLR+ NCAN	-1.23	-2.22	2.20E-11
2	CXCL12+ PAFAH1B1	1.32	-1.92	9.84 E-13
3	NCAN + SEZ6L	-0.08	-1.56	7.94 E-12
4	NCAN + TOMM40	-1.49	-1.34	1.75E-08
5	HAVCR1 + MLXIPL	-0.12	-1.20	2.06 E-05
6	LDLRAP1 + SORT1	-1.09	-1.04	6.78E-09
7	LDLR + LDLRAP1	-2.49	-0.92	2.36E-10
8	CELSR2 + LPL	-0.11	-0.86	8.7 E-05
9	MLXIPL + TOMM40	-1.94	-0.85	3.82 E-08
10	LDLR + MLXIPL	-2.26	-0.78	4.19E-11
11	HAVCR1+ SEZ6L	-0.67	-0.77	2.9 E-04
12	HAVCR1 + LDLRAP1	0.39	-0.73	5.40E-03
13	BCAM + LDLRAP1	-0.05	-0.70	5.20E-03
14	HAVCR1 + SORT1	-0.63	-0.58	1.98E-03
15	BCAM + HAVCR1	-0.51	-0.54	6.23E-02
16	MYBPHL + SIK3	2.30	0.96	7.91E-03
17	APOB + HMGCR	3.33	0.97	1.53E-03
18	SORT1 + TOMM40	-2.48	1.77	0.00E00
19	PAFAH1B1+ SIK3	3.30	2.44	8.37E-12
20	PCSK9 + TMEM57	3.21	2.44	3.74E-11
21	HAVCR1+LDLR	-2.24	-1.24	1.81E-10

Table 5. Validated gene interactions.

21 gene interactions were validated with liquid-phase transfection. Shown are the interaction effects of the forward transfection experiments, as well as the fdr-corrected p-values.

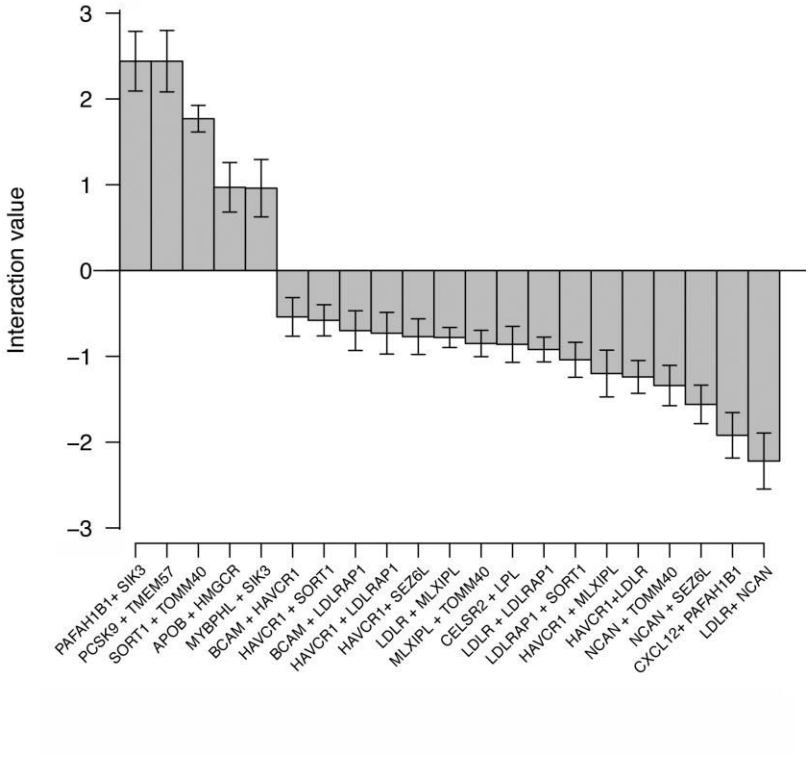


Figure 18. Validated gene interactions.

21 gene interactions were validated with liquid-phase transfection. The y-axis represents the interaction effects of the forward transfection experiments.

3.1.7.2 Relating Phenotypic Effect To Gene Expression

3.1.7.2.1 Evaluating Gene Expression

The use of a primary-tumor derived cell line, such as HeLa Kyoto, for RNAi screening raises the concern if the genes targeted are expressed and if their expression levels are reduced through RNAi. To address this question, HeLa Kyoto transcriptome RNA-sequencing data from the Simpson et al. secretion screen (Landry et al. 2013; Simpson et al. 2012) were explored, to assess if the genes tested with the co-RNAi screen are expressed in HeLa Kyoto cells. Moreover, qRT-PCR experiments were performed for 22 of the 30 genes screened with co-RNAi, and C_T values from qRT-PCR were juxtaposed to RNA-Seq coverage values, considering that both measurements reflect the mRNA level. The correlation between the two methods was strong (R=-0.63, p-value= 0.008354) (see Fig. 19).

Gene	RNA-seq (cov)	qRT-PCR (C_T)	knockdown detectable
GAPDH	2627.17	14.0	+
TOMM40	142.26	19.0	+
BCL7B	52.53	21.2	+
PAFAH1B1	39.47	20.2	+
PCSK9	34.98	19.5	+
LDLR	32.38	20.5	+
CELSR2	30.49	22.3	+
PVRL2	21.55	21.4	+
LDLRAP1	18.33	23.1	+
SIK3	16.34	23.9	+
LPL	15.78	21.4	+
TMEM57	15.16	22.1	+
HMGCR	12.18	20.3	+
SORT1	10.31	22.5	+
MLXIPL	5.85	24.8	+
BCAM	3.57	27.3	+
APOE	1.16	28.9	-
APOB	n.d.	36.0	-
CXCL12	n.d.	36.2	-
HAVCR1	n.d.	33.2	-
MYBPHL	n.d.	32.9	-
NCAN	n.d.	33.2	-
SEZ6L	n.d.	33.1	-

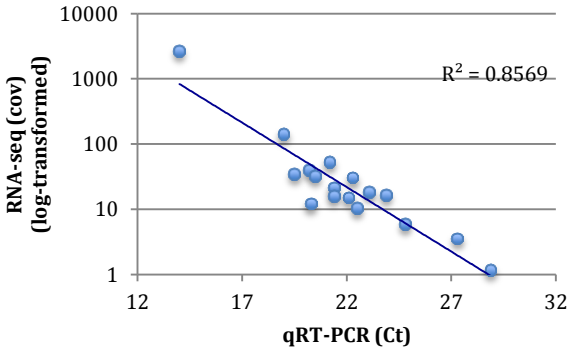


Figure 19. Correlation of RNA-Seq with qRT-PCR gene expression experiments

Both qRT-PCR and RNA-Seq experiments were utilized to quantify the mRNA levels of genes in HeLa Kyoto cells, by measuring the threshold cycle (Ct) or the coverage (cov), respectively. For 17 of 23 genes, for which Ct of qRT-PCR was below 29, RNA-Seq detected the mRNA and the knockdown could be measured with qRT-PCR. The correlation between the two methods was strong (Pearson's correlation coefficient R= -0.63, p-value: 0.008354) (n.d.: not detected).

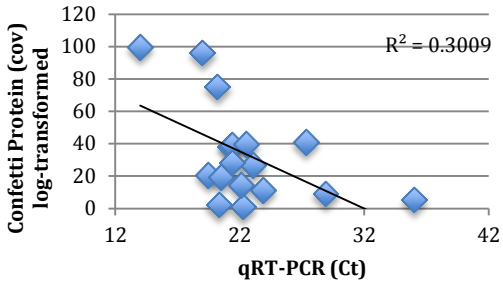
In agreement with the RNA-Seq data, gene knockdown could only be measured with qRT-PCR in genes with C_T<29 and that were detected with RNA-Seq. Out of 30 genes tested in the co-RNAi screen, 22 (73%) were detectable with RNA-Seq in the transcriptome of HeLa Kyoto cells. However, some of the genes that scored as interactors in the co-RNAi screen were neither detected by RNA-seq, nor by qRT-PCR.

With a view to investigating if the protein expression of such genes, that were not detectable at mRNA level, could be observed, the Confetti online resource of HeLa proteome (https://proteomics.swmed.edu/confetti/) was employed. This is a map of the HeLa proteome that was generated by digesting the HeLa lysate with a multiprotease approach, and subsequently analyzing the lysate digest with mass spectrometry (Guo et al. 2014). This approach led to the identification of more than 7,500 proteins.

Out of the 30 genes tested in the co-RNAi screen, 21 (70%) were detected in the proteome, according to the Confetti map. Noteworthy one gene that was not detected by RNA-seq and its knockdown could not be detected with qRT-PCR (*APOB*), as well as one gene that was detected by RNA-Seq, but its knockdown could not be detected (*APOE*), were detected in the proteome. Moreover, one gene that was detected with RNA-Seq and its knockdown could be measured, was not detected at the proteome level (*MLXIPL*). The correlation between the protein levels from Confetti and the mRNA levels from RNA-Seq was strong (R=0.72, p-value= 0.001055), but the correlation between the protein levels and the mRNA levels from qRT-PCR was weak (R= 0.327208) (see Fig. 20).

Both the transcriptome and the proteome resources were useful to provide a priori knowledge concerning gene expression levels at the mRNA and protein level in HeLa cells. However, neither of them was utilized with the aim to preclude genes from the co-RNAi screen. As was explained before, the expression of genes undetected by RNA-seq or proteomics might be lower than the detection threshold of these techniques.

Gene	RNA-seq (cov)	qRT-PCR (C_T)	Confetti (cov)
GAPDH	2627.17	14	99.7
TOMM40	142.26	19	96.12
BCL7B	52.53	21.2	38.12
PAFAH1B1	39.47	20.2	75.12
PCSK9	34.98	19.5	20.52
LDLR	32.38	20.5	19.42
CELSR2	30.49	22.3	1.03
PVRL2	21.55	21.4	39.22
LDLRAP1	18.33	23.1	26.62
SIK3	16.34	23.9	11.01
LPL	15.78	21.4	28.21
TMEM57	15.16	22.1	14.46
HMGCR	12.18	20.3	2.48
SORT1	10.31	22.5	39.71
MLXIPL	5.85	24.8	n.d.
BCAM	3.57	27.3	40.92
APOE	1.16	28.9	9.15
APOB	n.d.	36	5.28
CXCL12	n.d.	36.2	n.d.
HAVCR1	n.d.	33.2	n.d.
MYBPHL	n.d.	32.9	n.d.
NCAN	n.d.	33.2	n.d.
SEZ6L	n.d.	33.1	n.d.



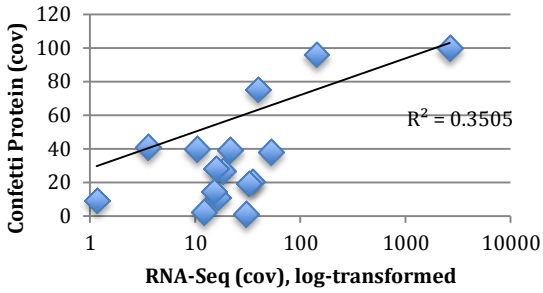


Figure 20. Correlation of RNA-Seq and qRT-PCR gene expression experiments with protein expression

For 17 out of 23 genes, for which the Ct of qRT-PCR was below 29, the protein was detected with the Confetti approach. Noteworthy one gene (MLXIPL) with Ct below 29 (Ct=24.8) the protein was not detected. The correlation between RNA-Seq and protein measurements was strong (Pearson's correlation coefficient R=0.72, p-value=0.001055), but the correlation between qRT-PCR and protein values was weak (Pearson's correlation coefficient R=0.327208).

3.1.7.2.2 ASSESSING KNOCKDOWN EFFICIENCY

In an attempt to correlate the observed effect in the cellular assay with the RNAi-mediated knockdown efficiency, RT-qPCR was utilized to quantify the mRNA levels in the cell lysate of the cells transfected with siRNA targeting the genes under study. In detail, knockdowns for all the 19 validated interactors of the screen were performed. For six (*APOB*, *CXCL12*, *HAVCR1*, *MYBPHL*, *NCAN*, *SEZ6L*) of the 19 genes tested, mRNA levels were too low to be detected and therefore knockdown could not be assessed. For the remaining 13 genes, the average knockdown was 80%. For 11 of the 13 genes, the knockdown ranged between 67% and 97%, while for the other 2 genes (*MLXIPL*, *TMEM57*) the knockdown ranged between 35% and 45% (see Tbl. 6). The anti-correlation of the knockdown levels with the functional effects observed was moderate (Pearson's correlation coefficient R= -0.42). For instance, strong phenotypic effects were observed for genes with a mild knockdown (<45%), such as *MLXIPL* and *TMEM57*. Similarly, very mild phenotypic effects were observed for genes with a strong knockdown (>95%), such as *PAFAH1B1* and *BCAM*.

siRNA treatment	% gene kd	effect on LDL uptake
BCAM + control	95.7	-0.16
CELSR2 + control	85.0	0.84
HMGCR + control	67.4	1.70
LDLR + control	76.8	-1.43
LDLRAP1 + control	92.9	1.11
LPL + control	96.7	-0.67
MLXIPL + control	35.6	0.89
PAFAH1B1 + control	95.3	0.37
PCSK9 + control	85.1	0.70
SIK3 + control	78.9	0.39
SORT1 + control	93.9	-0.56
TMEM57 + control	44.8	1.01
TOMM40 + control	88.6	-1.51

Table 6. Correlation of the gene knockdown with the functional effects on LDL uptake
The gene knockdown efficiencies are calculated after normalization to the GAPDH. The effect on LDL uptake shown is the median robust Z score of endocytosed DiI-LDL. The correlation of the gene knockdown levels with the functional effects was moderate (Pearson's correlation coefficient R= -0.42).

3.1.7.3 Effect of Gene Interactions on LDLR mRNA and protein level

Next, I asked whether the effects of the identified gene interactions on LDL-uptake are due to alterations of the LDLR expression at the transcriptional and/ or post-transcriptional level. Given the fact that LDLR is the key regulator of LDL endocytosis, an increase in LDLR expression should lead to increased LDL uptake and *vice versa*. In order to address this question, LDLR mRNA and protein levels upon knockdowns were assessed for the validated hits of the screen. Altogether, 20 double knockdowns for the interacting gene pairs as well as the corresponding single knockdowns were performed, and their effect on LDLR mRNA and protein levels was measured using qRT-PCR and Western blotting, respectively (Figures 21 and 23).

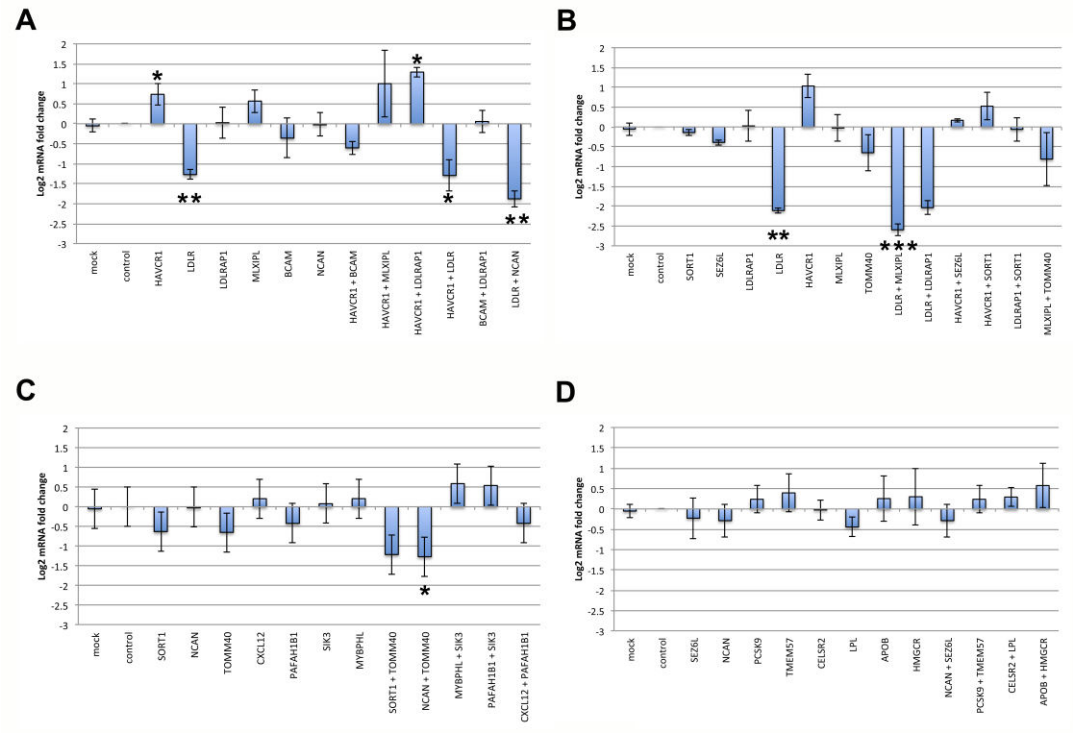


Figure 21. Effect of gene interactions on LDLR mRNA levels

Shown are mRNA levels of LDLR in HeLa Kyoto cells after 48 h knockdown of the identified interacting genes, transfected alone ($siRNA_A$+control siRNA) or in combination ($siRNA_A$+$siRNA_B$), after normalization to the control siRNA. The LDLR levels were normalized to the housekeeping gene GAPDH. The error bars represent the standard error of the mean (n=4, statistical significance compared to the control in two-sided Student´s t-test is indicated; *P<0.1; **P<0.01; ***P<0.001). (A-D) The siRNA concentrations used are shown in Tbl. S10 and numeric data in Tbl. S13.

A significant decrease in LDLR mRNA levels was observed upon the single knockdown of LDLR (at different siRNA concentrations), as well as for the double knockdowns LDLR+HAVCR1, LDLR+MLXIPL, LDLR+NCAN and NCAN+TOMM40.

In order to identify which of the 20 validated gene interactions are reflected on LDLR mRNA level, the additive effect of two single knockdowns was compared to the combined knockdown effect, by fitting a linear regression model (see section 7.4.7), as was done in the analysis of the co-RNAi screen. The interaction effect was calculated, and the threshold for the p-value was set at $pVal<10^{-1}$. With this analysis, five gene pairs (HAVCR1+LDLR, LDLR+MLXIPL, HAVCR1+SEZ6L, NCAN+TOMM40, LDLR+NCAN) were identified as interacting at the level of LDLR mRNA expression (see Fig. 22). The results of the linear regression analysis are shown in Tbl. S16.

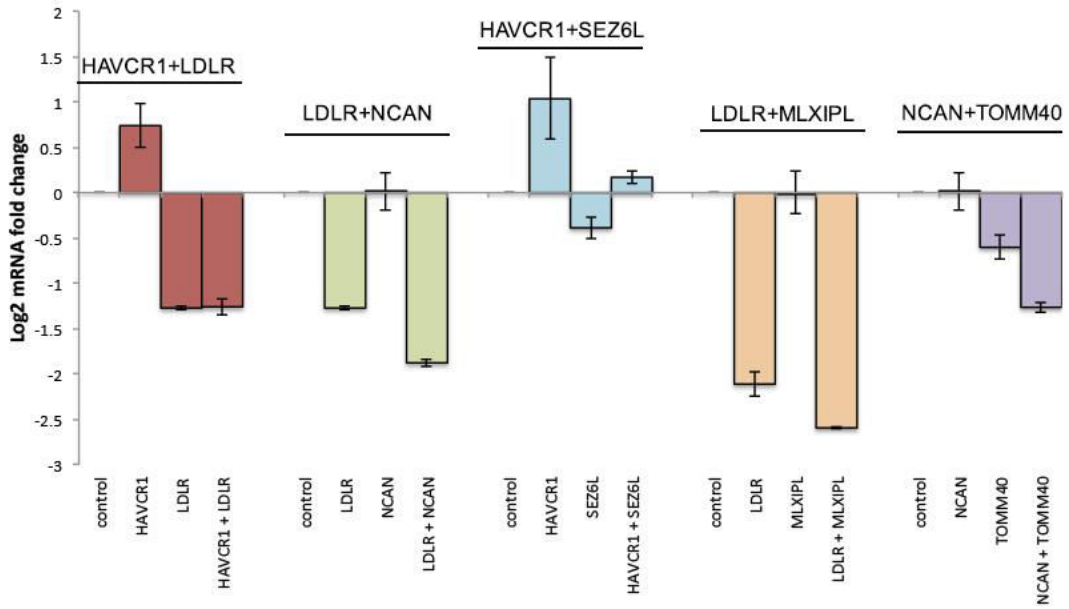


Figure 22. Gene interactions affecting LDLR mRNA levels.
Shown are the 5 gene pairs that were identified as interacting at the level of LDLR mRNA, by comparing the additive effect of the two single knockdowns to the effect of the double knockdown, using a linear regression model. For each gene pair, the mean (log2-transformed) fold change of mRNA of the control, the two single knockdowns as well as the double knockdown are shown. The LDLR levels were normalized to the housekeeping gene GAPDH. The error bars represent the standard error of the mean (n=4). The siRNA concentrations used are shown in Tbl. S10 and numeric data in Tbl. S11.

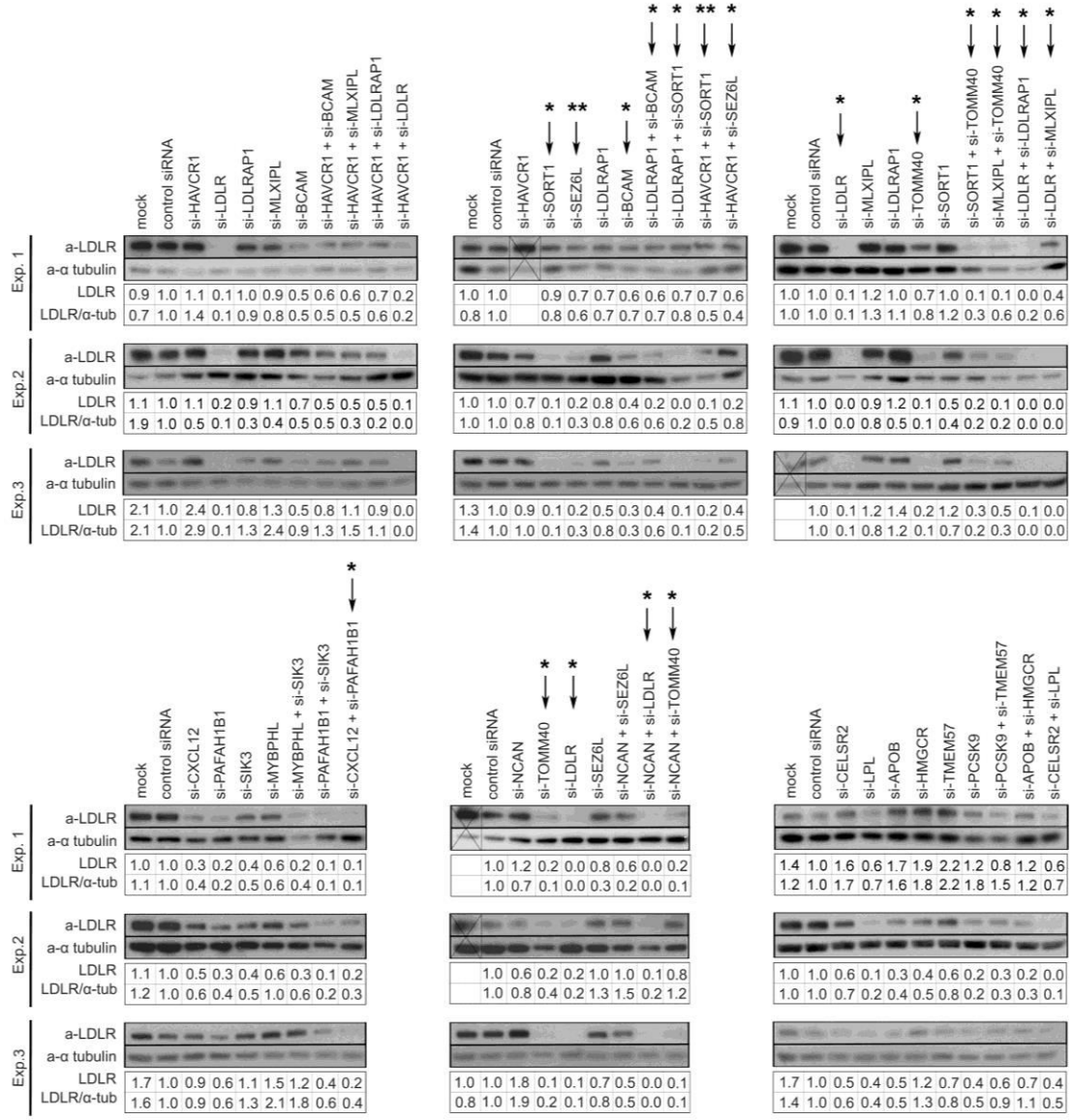


Figure 23. Effect of gene interactions on LDLR protein level.

LDLR protein levels in HeLa Kyoto cells were measured by Western blot upon 48 h knockdown following starvation of the cells (see section 7.1.6). The identified interacting genes were either transfected alone (siRNA$_A$+control siRNA) or in combination (siRNA$_A$+siRNA$_B$). LDLR was normalized to the housekeeping gene a-tubulin, after normalization of both to the control siRNA. Shown are results from three biological replicates. Arrows denote siRNAs that significantly (*<0.01; **<0.001, ***<0.0001) altered LDLR protein levels. Bands from lysates wherefrom no reliable results could be obtained were crossed out and excluded from the analysis. The siRNA concentrations used are shown in Tbl. S10 and numeric data in Tbl. S13.

A significant decrease in LDLR protein levels was observed upon single knockdown of *LDLR*, *TOMM40*, *SEZ6L*, *SORT1* and *BCAM*, as well as for the double knockdowns *LDLRAP1+BCAM*, LDLRAP1+SORT1, *HAVCR1+SORT1*, *HAVCR1+SEZ6L*, *SORT1+TOMM40*, *MLXIPL+TOMM40*, *LDLR+LDLRAP1*, *LDLR+MLXIPL*, *CXCL12+PAFAH1B1*, *NCAN+LDLR*, *NCAN+TOMM40*.

In order to identify which of the 20 validated gene interactions are reflected on LDLR protein level, the additive effect of two single knockdowns was compared to the combined knockdown effect, by fitting a linear regression model, as before (see section 7.4.7). The interaction effect was calculated, and the threshold for the p-value was set at $pVal<10^{-1}$. With this analysis, five gene pairs (*LDLR+LDLRAP1*, *MLXIPL+TOMM40, BCAM+LDLRAP1*, *HAVCR1+LDLR*, and *CXCL12+PAFAH1B1*,) were identified as interacting at the level of *LDLR* protein expression (see Fig. 24). The results of the linear regression analysis are shown in Tbl. S16.

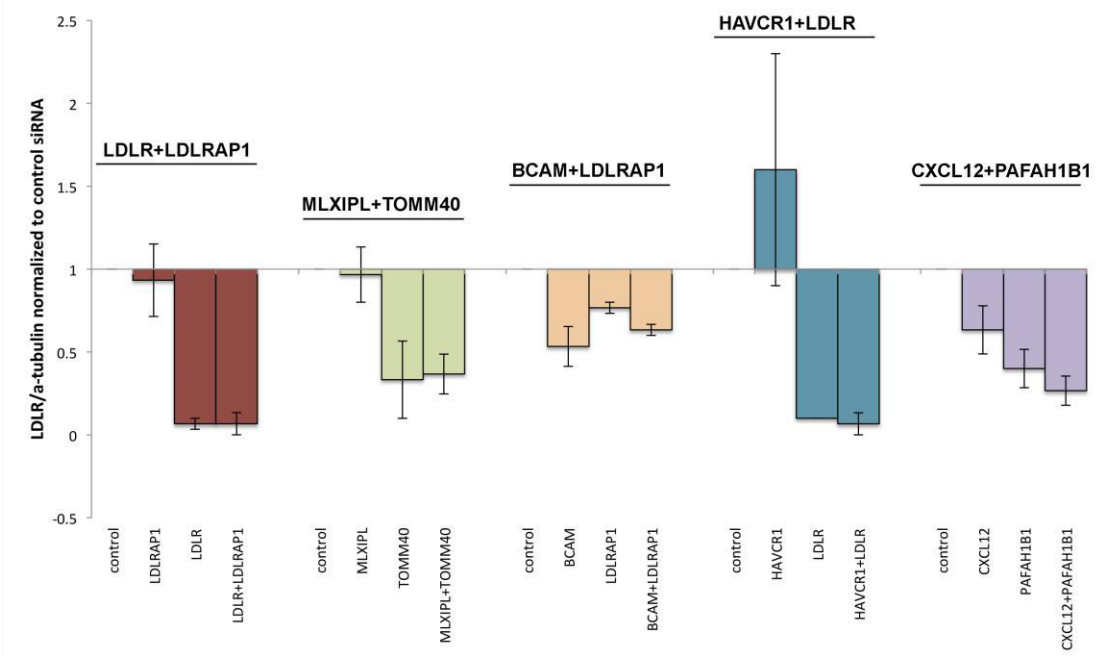


Figure 24. Gene interactions affecting LDLR protein levels.
Shown are the five gene pairs that were identified as interacting at the level of LDLR protein, by comparing the additive effect of the two single knockdowns to that of the double knockdown, using a linear regression model. For each gene pair, the median LDLR after normalization to a-tubulin is shown for the control siRNA, the two single knockdowns and the double knockdown. The error bars represent the standard error of the mean (SEM). The siRNA concentrations used are shown in Tbl. S10 and numeric data in Tbl. S12.

3.1.7.4 EFFECT OF KNOCKDOWNS ON SREBF MRNA LEVEL

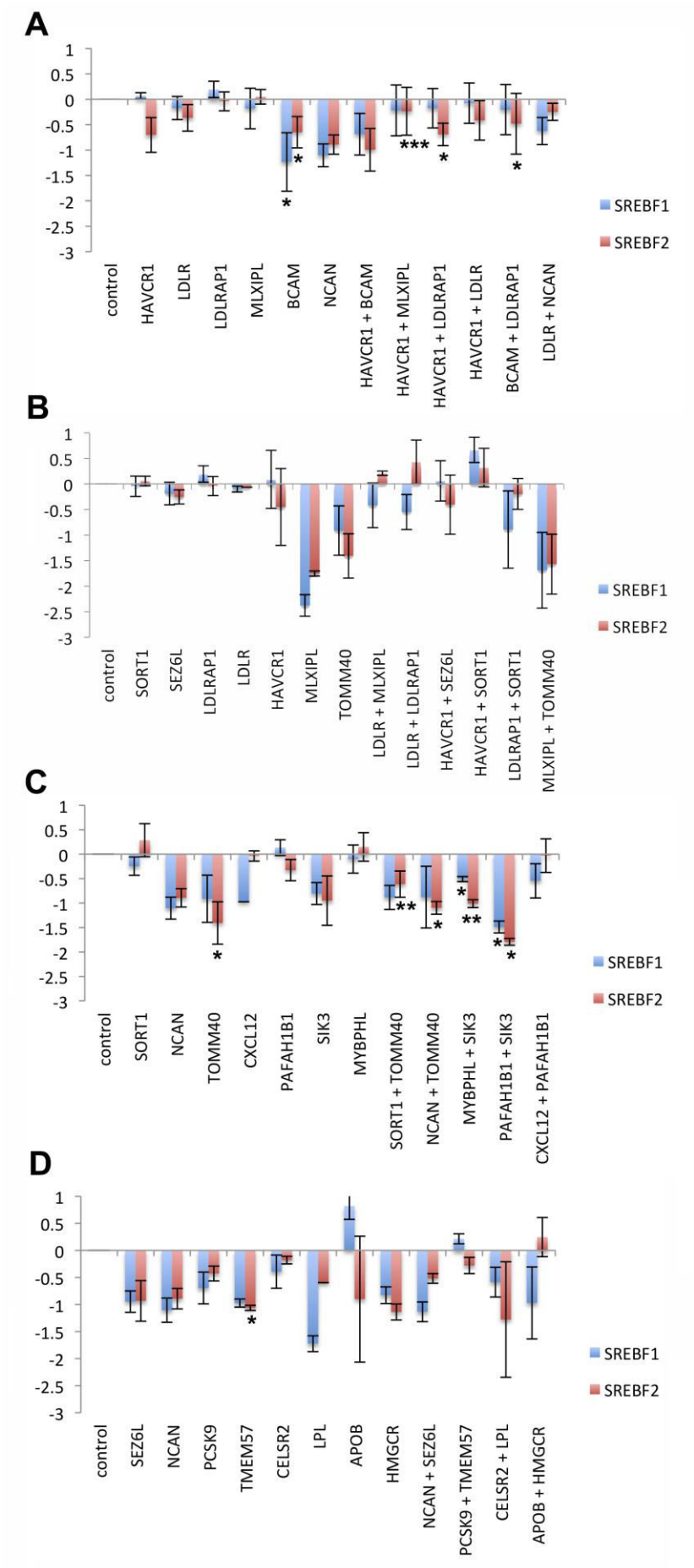
As was described (see section 1.3.2.2), LDLR expression is regulated by the SREBF transcription factors, in particular SREBF-1a and SREBF-2. In order to determine if the effect of the identified gene interactions on LDL uptake was due to an effect on the expression of *SREBF-1a* and/or *SREBF-2*, the mRNA levels of both genes upon knockdowns were assessed for the validated hits of the screen. Namely, 20 double knockdowns for the interacting gene pairs as well as the corresponding single knockdowns were performed, and their potential effect on *SREBF* mRNA levels was measured using qRT-PCR.

A significant decrease in SREBF1 levels was observed upon single knockdown of *BCAM*, as well as double knockdowns *MYBPHL+SIK3* and *PAFAH1B1+SIK3* (Fig. 25). Regarding SREBF2, a significant decrease in mRNA levels was observed upon single knockdown of BCAM, as well as double knockdowns *SORT1+TOMM40*, *NCAN+TOMM40*, *MYBPHL+SIK3*, *PAFAH1B1+SIK3*, *HAVCR1+MLXIPL*, *HAVCR1+LDLRAP1* and *BCAM+LDLRAP1* (see Fig. 25).

In order to identify which of the 20 validated gene interactions are reflected on the *SREBF1* and/or *SREBF2* mRNA level, the additive effect of two single knockdowns was compared to the combined knockdown effect, by fitting a linear regression model (see section 7.4.7). The interaction effect was calculated, and the threshold for the p-value was set at $pVal<10^{-1}$. With this analysis, four gene pairs were identified to interact at the *SREBF1* mRNA level; *SORT1+LDLRAP1*, *PCSK9+TMEM57*, *HAVCR1+SORT1* and *CELSR2+LPL* (see Fig. 26, for numerical data see Tbl. S17). At the mRNA level of SREBF2, eight interactions were identified; *PCSK9+TMEM57*, *LDLRAP1+SORT1*, *NCAN+TOMM40*, *LDLR+NCAN*, *LDLR+LDLRAP1*, *SORT1+TOMM40*, *MYBPHL+SIK3* and *LDLR+MLXIPL* (see Fig. 27, Tbl. S17).

Figure 25. Effect of gene interactions on SREBF1 and SREBF2 levels

Shown are mRNA levels of SREBF1 (blue) and SREBF2 (red) in HeLa Kyoto cells after 48 h knockdown of the identified interacting genes, transfected alone ($siRNA_A$+control siRNA) or in combination ($siRNA_A$+$siRNA_B$), after normalization to the control siRNA. The LDLR levels were normalized to the housekeeping gene GAPDH. The error bars represent the SEM (n=3, statistical significance compared to control in two-sided Student´s t-test is indicated; *P<0.1; **P<0.01; ***P<0.001). (A-D) The siRNA concentrations used are shown in Tbl. S10 and numeric data in Tbl. S15.


A
1
0.5
0
-0.5
-1
-1.5
-2
-2.5
-3
control
HAVCR1
LDLR
LDLRAP1
MLXIPL
BCAM
NCAN
HAVCR1 + BCAM
HAVCR1 + MLXIPL
HAVCR1 + LDLRAP1
HAVCR1 + LDLR
BCAM + LDLRAP1
LDLR + NCAN
SREBF1
SREBF2
B
control
SORT1
SEZ6L
LDLRAP1
LDLR
HAVCR1
MLXIPL
TOMM40
LDLR + MLXIPL
LDLR + LDLRAP1
HAVCR1 + SEZ6L
HAVCR1 + SORT1
LDLRAP1 + SORT1
MLXIPL + TOMM40
SREBF1
SREBF2
C
control
SORT1
NCAN
TOMM40
CXCL12
PAFAH1B1
SIK3
MYBPHL
SORT1 + TOMM40
NCAN + TOMM40
MYBPHL + SIK3
PAFAH1B1 + SIK3
CXCL12 + PAFAH1B1
SREBF1
SREBF2
D
control
SEZ6L
NCAN
PCSK9
TMEM57
CELSR2
LPL
APOB
HMGCR
NCAN + SEZ6L
PCSK9 + TMEM57
CELSR2 + LPL
APOB + HMGCR
SREBF1
SREBF2

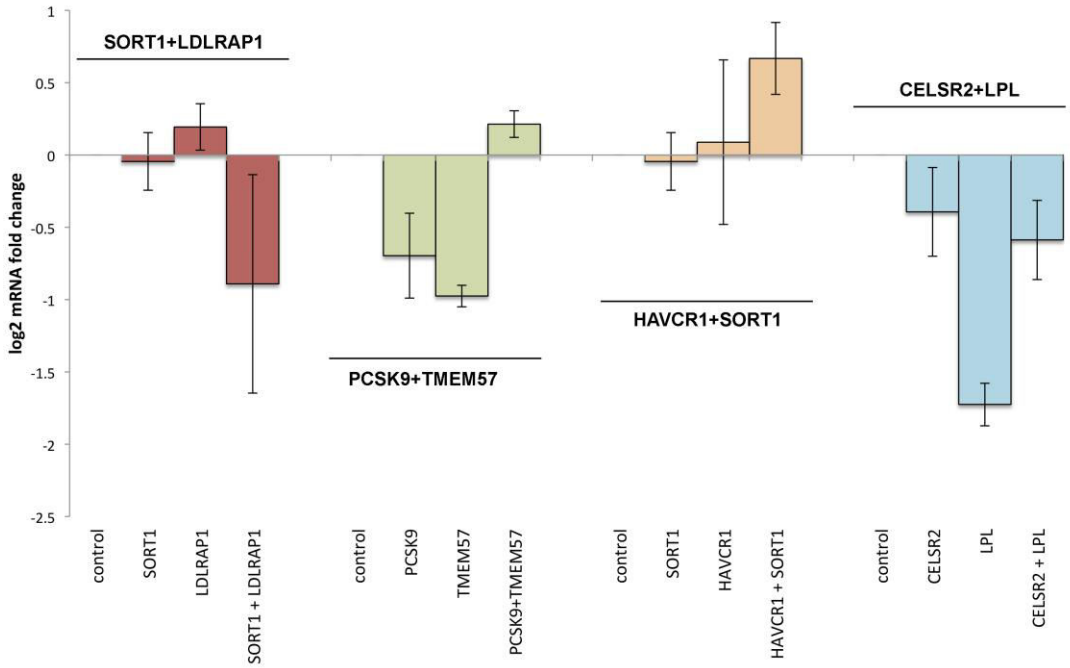


Figure 26. Gene interactions affecting SREBF1 mRNA levels.

Shown are the four gene pairs that were identified as interacting at the level of *SREBF1* mRNA, by comparing the additive effect of the two single knockdowns to the effect of the double knockdown, using a linear regression model. For each gene pair, the mean (log2-transformed) fold change of mRNA of the control, the two single knockdowns as well as the double knockdown are shown. The *SREBF1* levels were normalized to the housekeeping gene *GAPDH*. The error bars represent the standard error of the mean (n=4). The siRNA concentrations used are shown in Tbl. S10 and numeric data in Tbl. S.14

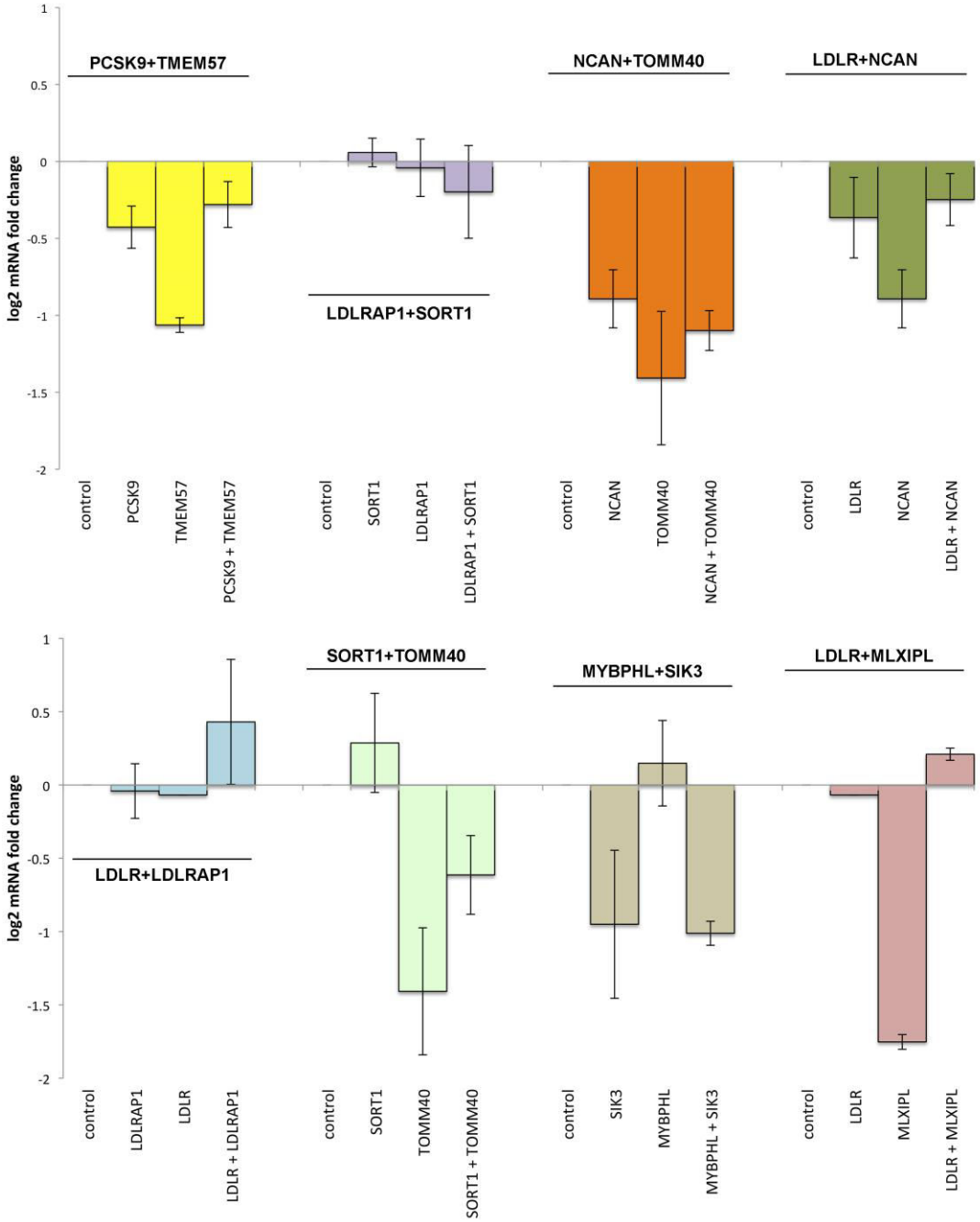


Figure 27. Gene interactions affecting SREBF2 mRNA levels.

Shown are the eight gene pairs that were identified as interacting at the level of *SREBF2* mRNA, by comparing the additive effect of the two single knockdowns to the effect of the double knockdown, using a linear regression model. For each gene pair, the mean (log2-transformed) fold change of mRNA of the control, the two single knockdowns as well as the double knockdown are shown. The *SREBF1* levels were normalized to the housekeeping gene *GAPDH*. The error bars represent the standard error of the mean (n=4). The siRNA concentrations used are shown in Tbl. S10 and numeric data in Tbl. S.14.

For 21 gene pairs in total, quantitative data for LDL uptake, LDLR mRNA and protein levels, as well as SREBF1 and SREBF2 mRNA levels are available. The most interesting gene pairs from this analysis are provided in Tbl. 7 (all results are available in Tbl. S18). The findings of these experiments were more complicated as it would have been expected. For example, *PCSK9+TMEM57* double knockdown resulted in increased *SREBF1* mRNA but decreased *SREBF2* mRNA. Another unexpected example is *LDLR+NCAN*, which decreased LDL uptake and LDLR mRNA levels, but also decreased *SREBF2* mRNA, while the contrary would be expected. The same was true for *NCAN+TOMM40*.

Table 7. Combined results from secondary experiments and co-RNAi for 10 selected gene pairs.
Shown is the effect on LDL uptake, LDLR mRNA and protein, as well as SREBF1 and SREBF2 mRNA, upon double knockdowns of the genes which were identified by the screen as interacting. Interaction effects are highlighted with pink (increased) or green (decreased).

		LDL	LDLR mRNA	LDLR protein	SREBF1	SREBF2		
	Treatment	**Robust Z score**	**log2 (fold change)**	**log2(LDLR/a-tubulin)**	**log2 (fold change)**	**log2 (fold change)**	**correlation**	**anti-correlation**
1	LDLR+ NCAN	**-1.23**	**-1.88**	-3.91	-0.63	**-0.25**	LDL-LDLR mRNA-SREBF2 mRNA	
2	NCAN + TOMM40	**-1.49**	**-1.27**	-1.10	-0.88	**-1.10**	LDL-LDLR mRNA-SREBF2 mRNA	
3	LDLRAP1 + SORT1	**-1.09**	-0.06	-1.45	**-0.89**	**-0.20**	LDL-SREBF1 mRNA -SREBF2 mRNA	
4	LDLR + LDLRAP1	**-2.49**	-2.04	**-3.91**	-0.55	**0.43**	LDL-LDLR protein	LDL-SREBF2 mRNA
5	MLXIPL + TOMM40	**-1.94**	-0.81	**-1.45**	-1.69	-1.57	LDL-LDLR protein	
6	LDLR + MLXIPL	**-2.26**	**-2.60**	-2.32	-0.42	**0.21**	LDL-LDLR mRNA	LDL-SREBF2 mRNA
7	MYBPHL + SIK3	**2.30**	0.59	-0.10	-0.51	**-1.01**		LDL-SREBF2 mRNA
8	SORT1 + TOMM40	**-2.48**	-1.21	-2.10	-0.89	**-0.61**	LDL-SREBF2 mRNA	
9	PCSK9 + TMEM57	**3.21**	0.24	-0.15	**0.21**	**-0.28**	LDL-SREBF1 mRNA	LDL-SREBF2 mRNA
10	HAVCR1+LDLR	**-2.24**	**-1.29**	**-3.91**	-0.08	-0.42	LDL-LDLR mRNA-LDLR protein	

3.1.8 CLUSTERING OF INTERACTIONS

In an effort to allocate the genes identified as interactors from the co-RNAi screen onto genetic pathways and networks controlling LDL-cholesterol regulation, the 21 validated GIs from the RNAi screen were classified into 3 gene-gene interaction categories; suppression, epistasis and synergy. The classification was done according to the definitions for these classes, which are quoted here from Perez et al. (Pérez-Pérez et al. 2009):

(i) Suppression; "when the double knockdown is closer to the wild-type condition than either of the single knockdowns"

(ii) Epistasis; "when the double knockdown resembles the phenotype of one of the singles, but not the other"

(iii) Synergy; "when the joint contribution to the phenotype made by both knockdowns is greater than the sum of their individual effects". The three interaction classes are shown in Fig. 28.

Out of 21 GIs, 10 were classified as "synergistic", 8 as "epistatic" and 3 as "suppressive" (see Tbl. 8).

Suppression	Epistasis	Synergy
HAVCR1 + MLXIPL	NCAN + SEZ6L	CXCL12+ PAFAH1B1
CELSR2 + LPL	NCAN + TOMM40	HAVCR1 + LDLRAP1
BCAM + LDLRAP1	LDLRAP1 + SORT1	HAVCR1 + SORT1
	LDLR + LDLRAP1	BCAM + HAVCR1
	MLXIPL + TOMM40	LDLR+ NCAN
	LDLR + MLXIPL	MYBPHL + SIK3
	HAVCR1+ SEZ6L	PAFAH1B1+ SIK3
	SORT1 + TOMM40	PCSK9 + TMEM57
		HAVCR1+LDLR
		BCAM + HAVCR1

Table 8. Classification of gene interactions.

Shown is the classification into the 3 described genetic interaction classes, of the 21 gene-gene interactions identified from the RNAi screen that were validated.

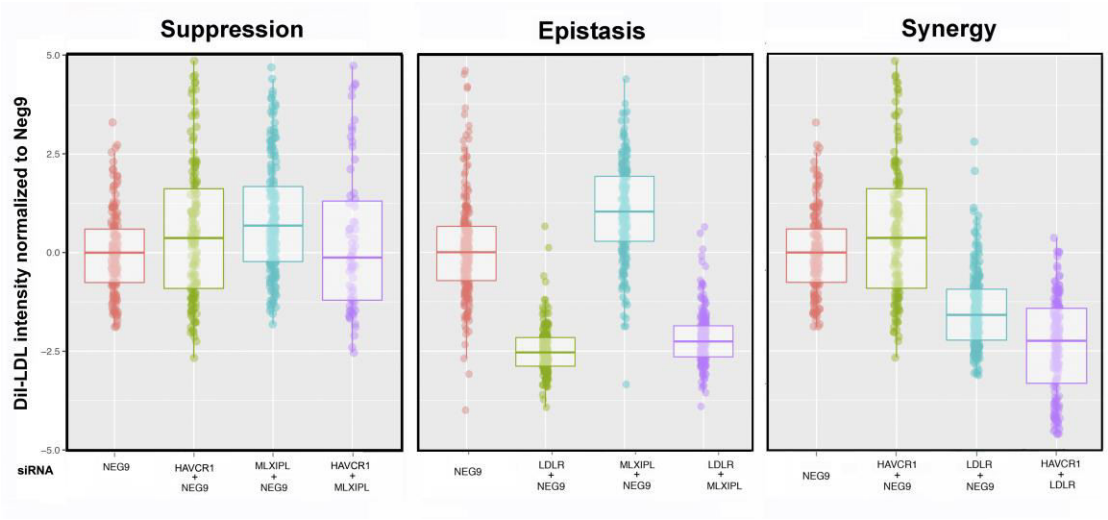


Figure 28. The three gene interaction classes identified in the co-RNAi screen.
Shown are representative examples of two interacting genes, for each of three identified classes of gene-gene interactions. The boxplots show the median intensity of internalized DiI-LDL, normalized to the control, for the two single knockdowns (transfected together with the control siRNA), as well as for the double knockdown, for each pair of genes.

3.1.9 Gene-Gene Interaction Model Network construction

With an aim to reveal potential functional relationships among the interacting genes identified with co-RNAi, a gene interaction network was constructed based on the 21 validated gene interactions identified. These interactions were overlapped with known interactions from the STRING, among all 30 genes tested in the co-RNAi screen (see Fig. 29). Out of the 30 genes tested, 24 genes were taking part in interactions, while six did not interact with any other gene. Out of the 24 interacting genes, 17 genes were connected into a main network (*APOB*, *APOE*, *BCAM*, *CELSR2*, *HAVCR1*, *HMGCR*, *LDLR*, *LDLRAP1*, *LPL*, *MLXIPL*, *MYLIP*, *NCAN*, *PCSK9*, *SEZ6L*, *SORT1*, *TMEM57* and *TOMM40*), five other genes were connected into a smaller network (*CXCL12*, *MYBPHL*, *PAFAH1B1*, *PAFAH1B2* and *SIK3*), and two other genes were only connected with each other (*WDR12* and *ZNF259*).

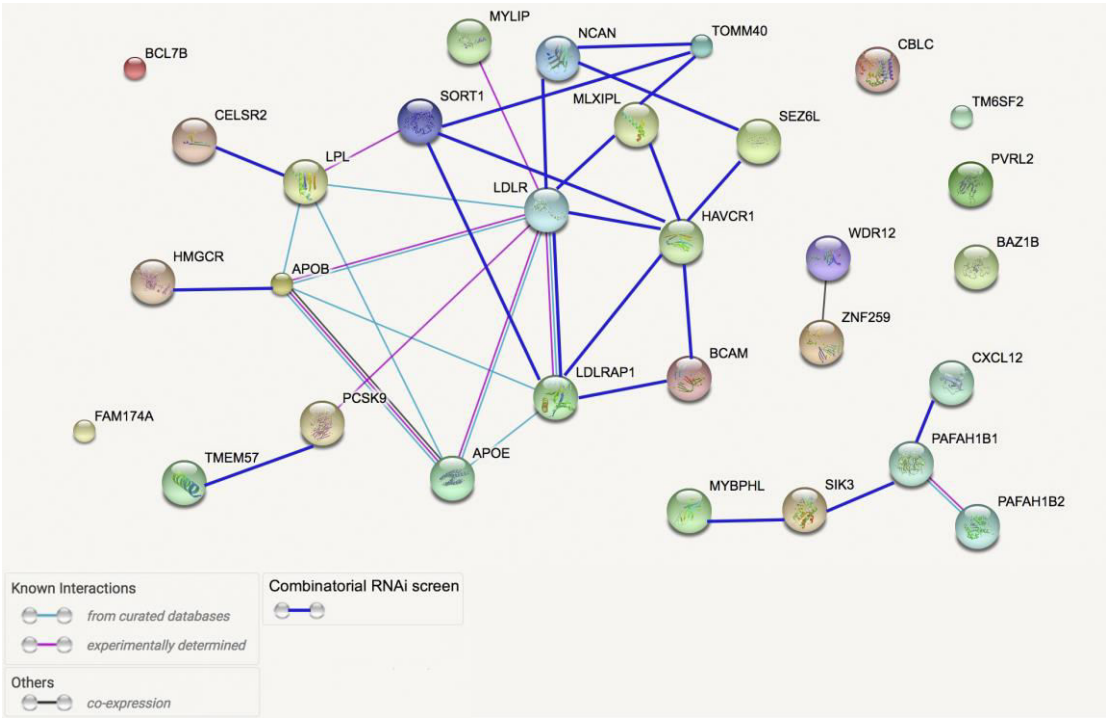


Figure 29. Model network of gene-gene interactions affecting LDL uptake.
In this network that was constructed using STRING, all the 30 genes tested in the co-RNAi screen are shown. All pairwise measurements between the 30 depicted genes have been tested, and interacting genes are depicted as connected with a dark blue line. The light blue and pink lines represent known interactions, whereas the black lines represent gene co-expression.

3.1.10 Network Analysis

In order to investigate the genetic interaction network that was constructed based on data from co-RNAi experiments, the large network was sub-divided into smaller sub-networks, such as the "triangles", that can be observed, which can then be further investigated. One such triangle is the network of *LDLR*, *HAVCR1* and *MLXIPL* (see Fig. 30). Within this sub-network, each gene interacts with the two other genes.

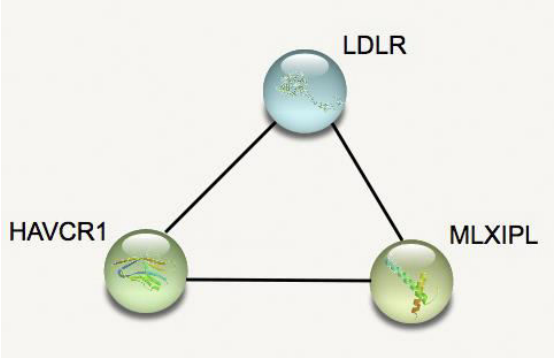


Figure 30. A sub-network of the GI model network.
LDLR, MLXIPL and HAVCR1 were shown to interact with each other in regulating LDL-uptake, therefore form a gene-interaction network that can be further investigated.

In this network, the knockdown of both *MLXIPL* and *HAVCR1* are potentiating the negative effect of *LDLR* knockdown on LDL uptake, upon co-depletion. On the other hand, the co-depletion of HAVCR1 and MLXIPL is reversing the positive effects of the two single knockdowns, leading to a down-regulation of LDL uptake. With these results in mind, I turned into the literature in order to construct hypotheses on how these 3 genes might interact in order to regulate the endocytosis of LDL. *LDLR* is a well-described gene that encodes for the receptor of LDL, and is responsible for its endocytosis, as was already described (see section 1.3.2.3). *HAVCR1* encodes a phosphatidylserine (PtdSer) receptor, which binds with its extracellular immunoglobulin variable domain (IgV domain) to PtdSer on the viral envelope of Hepatitis A virus and other viruses (s.a. EBOV), mediating their internalization. Moreover, it was shown that HAVCR1 can bind and internalize both oxidized and native LDL (Ichimura et al. 2008). Based on the above, I hypothesized that HAVCR1 might act as an alternative receptor for LDL, with lower affinity for LDL than LDLR, or alternatively might act as a chaperone for LDLR, thus promoting LDL uptake. *MLXIPL*, also known as *CHREBP* (Carbohydrate Response Element Binding Protein), encodes a

glucose-sensitive transcription factor that has been involved in lipogenesis (Cha and Repa 2007; Stoeckman, Ma, and Towle 2004). I hypothesized that MLXIPL might be involved in the transcriptional regulation of LDLR and/ or HAVCR1.

With a view to investigating these hypotheses, I performed some follow-up experiments for LDLR and HAVCR1.

3.2 Cell-Biological Characterization of Selected Screen Hits

3.2.1 HAVCR1 knockdown does not decrease LDL binding

First, in order to investigate the hypothesis that HAVCR1 might act as an alternative receptor for LDL, I asked if the depletion of HAVCR1 had an effect on the binding of LDL on the plasma membrane, and if its combined knockdown together with LDLR could potentiate the down-regulating effect of LDLR knockdown on LDL binding. To answer these questions, I depleted HAVCR1 and LDLR alone and in combination, using RNAi, and performed the LDL-binding assay (for details see section 7.1.5.2). The knockdown of HAVCR1 alone resulted to a significant increase on LDL binding, as compared to the control siRNA (TTEST; p=0.004199), whereas the combined knockdown of LDLR and HAVCR1 had no significant difference from the single knockdown of LDLR on LDL binding (see Fig. 31).

3.2.2 HAVCR1 overexpression does not rescue the effect of LDLR knockdown on LDL uptake

Next, I asked if HAVCR1 could function as a receptor for LDL upon depletion of LDLR. To test if HAVCR1 could rescue the effect of LDLR knockdown, LDLR was depleted using RNAi, and GFP-tagged HAVCR1 was overexpressed and the effect on LDL uptake was measured. The overexpression of HAVCR1 in cells that had been depleted of LDLR could not rescue the effect of LDLR knockdown (see Fig. 32).

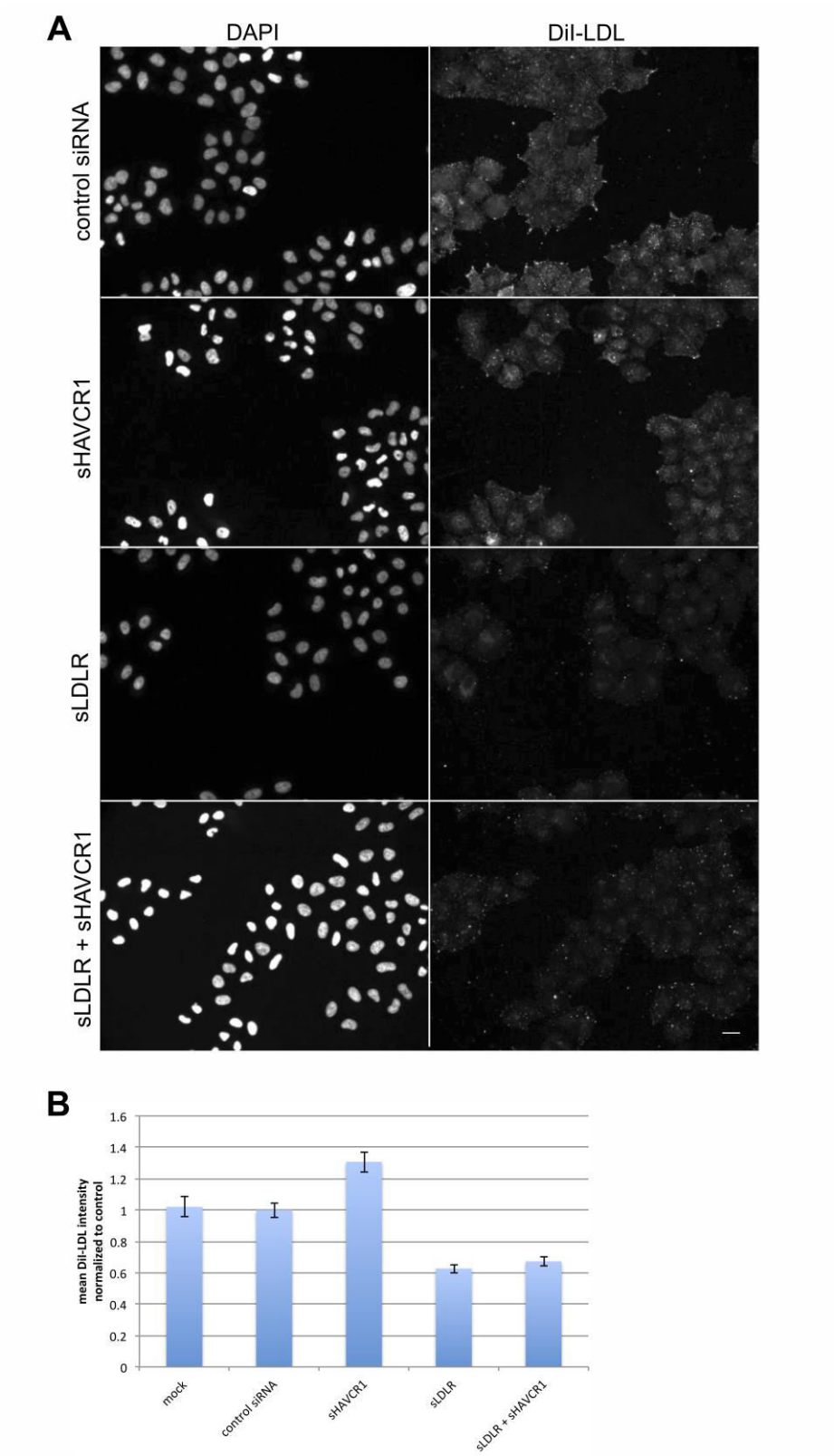


Figure 31. The effect of single and double knockdowns of LDLR and HAVCR1 on LDL binding and quantification.

LDLR knockdown significantly decreased LDL binding, whereas HAVCR1 slightly increased LDL binding. The combined knockdown of LDLR and HAVCR1 did not potentiate the effect of LDLR knockdown on LDL binding (scalebar=10 µm).

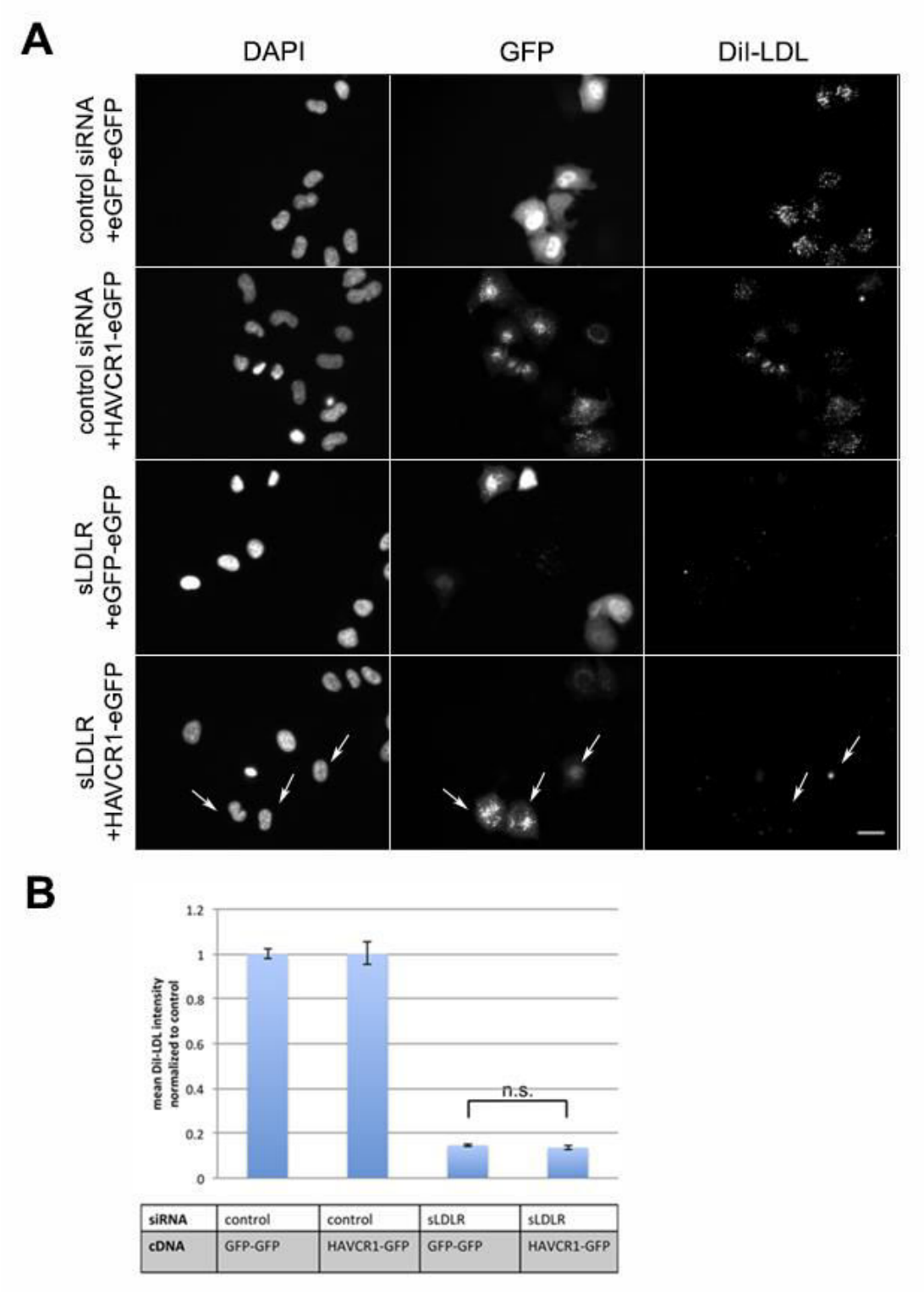


Figure 32. Rescue of LDLR knockdown with HAVCR1 overexpression.

HeLa Kyoto cells transfected with control siRNA or siRNA targeting LDLR, and with either GFP-GFP plasmid or HAVCR1-GFP. The LDL-uptake assay was performed to assess the ability of HAVCR1 to rescue the effect of LDLR knockdown on LDL uptake. (a) Representative cells for LDL-uptake assay (b) Quantification of the mean membrane-bound DiI-LDL, normalized to the control cells (scalebar=10 µm).

3.3 HAVCR1-RARE VARIANTS OVEREXPRESSION EFFECT ON LDL UPTAKE

3.3.1 VARIANT SELECTION

As a first step in investigating the role of rare variants of our interacting genes in CVD, I turned to one of the most prominent interactors of the screen, HAVCR1. As I hypothesize that HAVCR1 may play an important role in lipid regulation, I investigated the effect of a number of mutations of this gene on LDL uptake. For this purpose, 19 variants were extracted from the Exome Variant Server (http://evs.gs.washington.edu/EVS/), from the NHLBI Exome Sequencing Project (ESP). Of those 19 variants, 15 were rare (MAF<5%).

In general, ESP samples were selected to contain controls, the extremes of specific traits (LDL and blood pressure), and specific diseases (s.a. early onset MI and early onset stroke). The group includes some of the largest well-phenotyped populations in the United States, representing more than 200,000 individuals altogether from studies, such as the Framingham Heart Study (FHS), the Atherosclerosis Risk in Communities (ARIC), and the Multi-Ethnic Study of Atherosclerosis (MESA).

14 of these 19 variants were predicted to have a possibly damaging or probably damaging effect on protein function, 4 were predicted benign, and 1 had unknown effect, according to PolyPhen-2 (Polymorphism Phenotyping v2) (http://genetics.bwh.harvard.edu/pph2/) (see Tbl. 9).

The 19 variants were cloned into GFP expressing vector and expressed in HeLa Kyoto cells for 24 hours. The effect of the variants' overexpression on LDL uptake was measured.

rs ID	AA change	Alleles	MAF	PolyPhen2 Prediction
rs368474218	W47R	t139c	0.0244	probably damaging
rs373839023	V69I	g205a	0.0081	probably damaging
rs370454823	V69G	t206g	0.0081	probably damaging
rs56084311	D99H	g295c	0.8518	probably damaging
rs199849162	S100G	a298g	0.0241	possibly damaging
rs370980439	R110C	c328t	0.0081	probably damaging
rs377678930	N114S	a341g	0.0082	probably damaging
rs373938216	V138I	g412a	0.0078	possibly damaging
rs199816459	T149A	a445g	0.0078	possibly damaging
rs200642665	V156I	g446a	0.4005	benign
rs61734035	T174M	c521t	1.3038	probably damaging
rs1553316	L179P	t536c	14.6004	benign
rs12522248	T207A	a619g	24.6885	benign
rs369956191	E223stop	g667t	0.0081	unknown
rs200392856	D264N	g790a	0.0084	possibly damaging
rs376460912	Y297C	a890g	0.0083	probably damaging
rs370221161	K317E	a949g	0.0082	probably damaging
rs373996191	A343E	c1028a	0.0084	possibly damaging
rs141023871	M158T	t473c	42.7464	benign

Table 9. The 19 HAVCR1 variants extracted from the Exome Variant Server.
The rs (reference SNP cluster ID) number, the aminoacid substitution, the coded allele and the PolyPhen-2 prediction for each variant are shown.

3.4 RESULTS OF LDL UPTAKE ASSAY UPON OVEREXPRESSION OF HAVCR1 MUTATIONS

18 out of 19 variants had a significant inhibitory effect on LDL uptake upon overexpression (see Fig. 33 and Fig. 34). Interestingly, 4 variants that were predicted as benign by Polyphen-2 (188, 190, 191 and 230) had a significant inhibitory effect as well, while a variant that was predicted as probably damaging (186) had no significant effect.

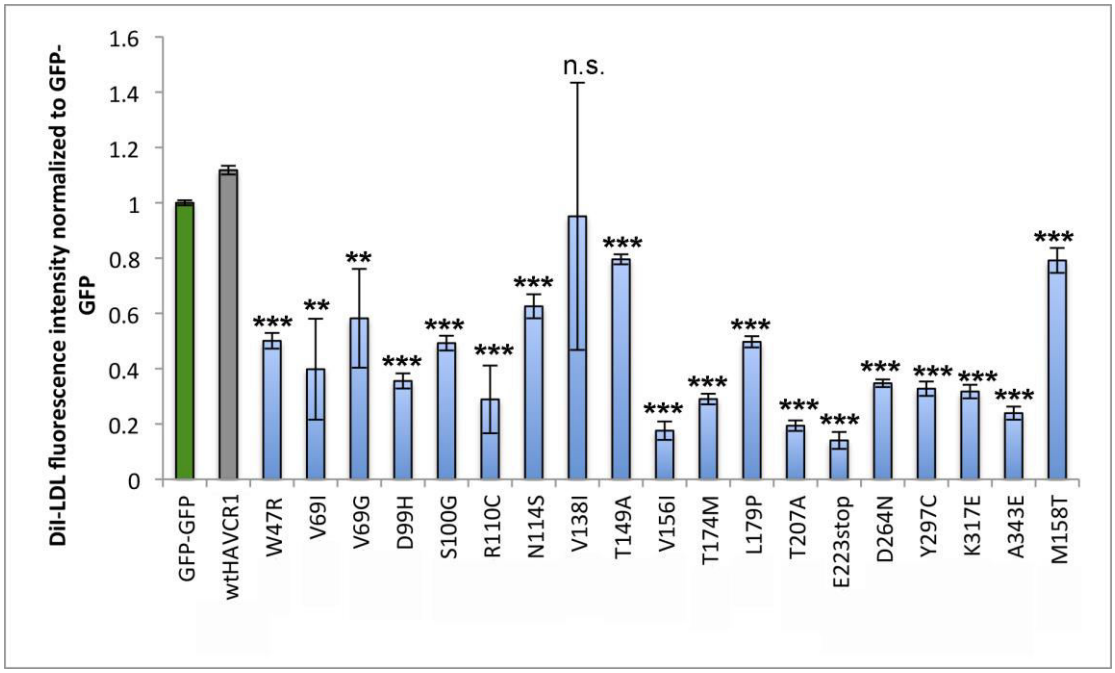


Figure 33. Effect of overexpression of GFP-tagged HAVCR1 variants in HeLa Kyoto cells.
The y-axis shows the fluorescence DiI-LDL intensity normalized to GFP vector-transfected cells. Blue: extracellular, pink: transmembrane, yellow: cytoplasmic (n.s.: not significant, * p<0.05, ** p<0.01, *** p<0.001, **** p<0.0001).

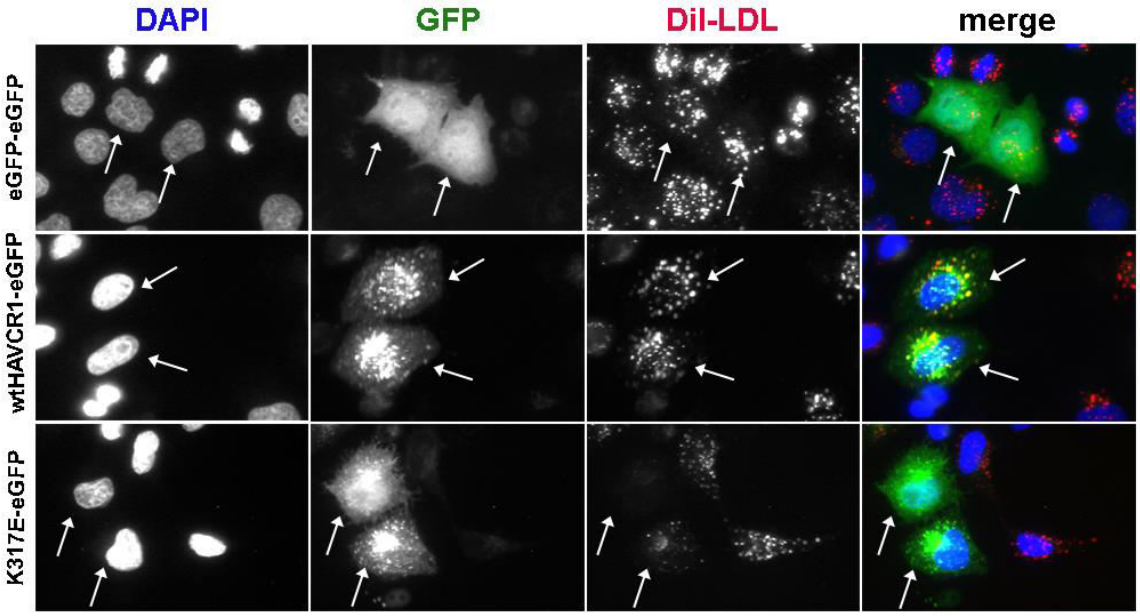


Figure 34. The effect of overexpression of wtHAVCR1-GFP and one tagged mutant on LDL uptake.
The first row shows cells transfected with GFP-vector without the protein. The second row shows cells transfected with the wild type HAVCR1 protein, and the third row shows cells transfected with one of the variants (K317E) that had a strong downregulating effect on LDL uptake.

3.5 SNP CO-OCCURRENCE DETECTION

For each of the genes in the GI network there have been found in GWA studies a number of SNPs associated with lipid traits and CVD. An interesting question that would give further insight into the molecular basis of CVD is: do SNPs of interacting genes co-occur? More precisely, can we detect SNPs from two or more interacting genes occuring in the same patient? Clinical data pointing to this direction will further support the *in vitro* results.

In order to address this question, the lead SNPs for each of the 30 genes that were tested in the co-RNAi screen were extracted from the PhenoScanner of the University of Cambridge (http://www.phenoscanner.medschl.cam.ac.uk/phenoscanner); a database holding publicly available results from large-scale genetic association studies. I noticed that the most significant lipid SNPs for the genes tested in the co-RNAi screen came from the Global Lipids Genomics consortium (Willer et al. 2013), the largest meta-GWAS study published to date. For comparisons, I decided to extract p-values from the same source. Therefore for lipid traits (LDL, HDL, TG, Tc) p-values for SNPs were only extracted from the above study, whereas for CAD, CHD and MI p-values were extracted from other studies in the PhenoScanner.

The total 144 lead SNPs extracted (Tbl. S21) are located on 14 genomic loci, where 28 of the 30 genes that were screened in this work here are mapped (*APOB*, *APOE*, *BAZ1B*, *BCAM*, *BCL7B*, *CBLC*, *CELSR2*, *HAVCR1*, *HMGCR*, *LDLR*, *LDLRAP1*, *LPL*, *MLXIPL*, *MYBPHL*, *MYLIP*, *NCAN*, *PCSK9*, *PVRL2*, *SIK3*, *SORT1*, *TM6SF2*, *TMEM57*, *TOMM40*, *ZNF259*). For two genes (*PAFAH1B2*, *SEZ6L*), SNPs have not yet reached statistical significance in GWAS.

The identified SNPs were tested for SNP-SNP interactions in a cohort of 4893 individuals from the Bioimage study, in collaboration with Heiko Runz (Merck Research Laboratories, US).

The Bioimage study is a multi-ethnic, observational study of the characteristics of subclinical atherosclerosis and cardiovascular disease, in a population of 7,300 US adults (55-80 years at baseline) at risk for, but without clinical cardiovascular disease (Muntendam et al. 2010). This study used measurements from imaging methods (including MRI for carotid and aortic plaques, coronary CT angiography, and PET/CT

for carotid and aortic inflammation), circulating biomarkers (s.a. blood lipids) and risk factors that predict progression to CVD (Natarajan et al. 2015).

Of the 144 SNPs that were extracted from the PhenoScanner, 70 were present on the Illumina Exome Array that was used for the genotyping in the Bioimage study (Muntendam et al. 2010), while 47 were not present on the Exome Chip but had proxies that were present on it, and 27 were neither present nor had proxies on the Exome Chip.

For the (70) SNPs that were present on the exome chip, as well as for the proxies of those not present (7 after removal of the identical ones), single-variant LDL association test was performed by Gulum Kosova, based on the allele-frequency of the LDL-raising allele in the Bioimage study. In order to test for pairwise SNP-SNP interactions, the following 4 models were tested against the null hypothesis model;

model 1: "snp1 only": $Y = \beta_0 + \beta_1 * snp1$

model 2: "snp2 only" : $Y = \beta_0 + \beta_2 * snp2$

model 3: "additive" : $Y = \beta_0 + \beta_1 * snp1 + \beta_2 * snp2$

model 4: "interaction": $Y = \beta_0 + \beta_1 * snp1 + \beta_2 * snp2 + \beta_3 * snp1 * snp2$,

to determine which model best explains the LDL phenotype.

The individual lipid phenotypes were provided in residuals, that is blood lipid phenotypes (LDL, HDL, TG), adjusted for covariates (s.a. age, BMI, alcohol intake, smoking, lipid-lowering treatment status) using linear regression models. Based on the individual lipid phenotypes, the number of individuals, as well as the mean residual LDL value was calculated for individuals carrying 0, 1, 2, 3 or 4 LDL-raising alleles (sum of snp1 and snp2 LDL-raising alleles).

With this analysis, none of the SNP pairs reached statistical significance for the "interaction" model. However, 36 SNP pairs reached statistical significance for the "additive" model (see Tbl. S22). Moreover, for these 36 pairs of SNPs it was shown that the LDL residuals increased with the number of LDL-increasing alleles. Namely, if one carries 3 or 4 risk alleles, one has higher chance of having high LDL blood levels, and as a consequence, higher risk for cardiovascular disease.

Interestingly, 3 pairs of SNPs (*LPL* + *CELSR2*, *APOB* + *HMGCR*, *LDLR*+*NCAN*) that were identified with this analysis to fit in the "additive" model correspond to genes that were identified as interactors for LDL-uptake in my co-RNAi screen (see Fig. 35).

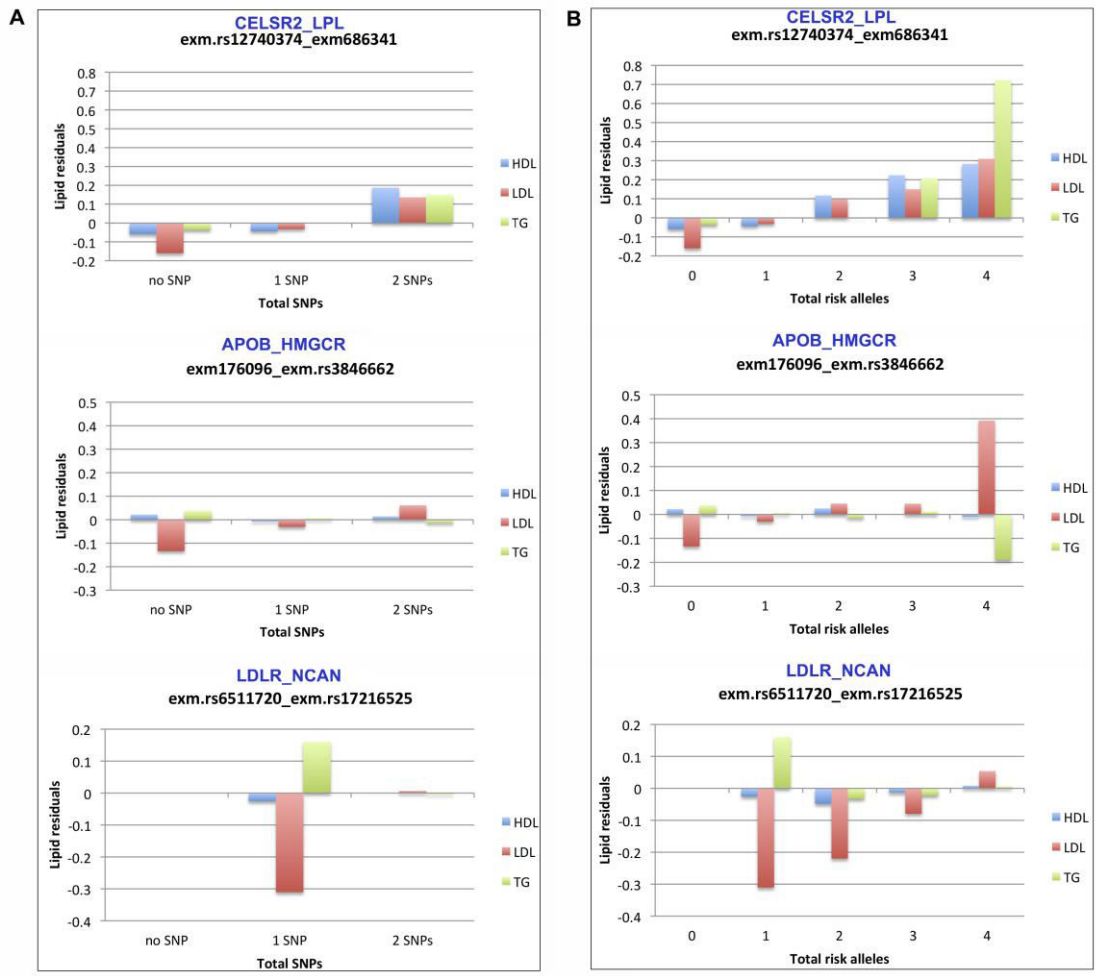


Figure 35. SNP co-occurrence effect on lipid levels for the 3 SNP pairs from Bioimage cohort that correspond to gene interactions from my co-RNAi screen

(A) The lipid levels (HDL, LDL, TG) in individuals where both SNPs are present (2 SNPs) versus individuals carrying 1 of the two SNPs (1 SNP) or no SNP for the three SNP pairs that were identified as having an additive effect in the Bioimage cohort and were interactors in my co-RNAi screen. (B) The lipid levels (residuals) for HDL, LDL and TG (y axis) are plotted against the total number of LDL-increasing alleles (x axis), in cases of SNP co-occurence (2, 3 or 4 risk alleles) and non co-occurence (0 or 1 risk alleles), for the three SNP pairs that correspond to genes that interacted on LDL uptake in my co-RNAi screen. The gene mapped closest to each SNP is shown above the SNP ID.

For these three SNP pairs, I extracted from the Bioimage cohort data the lipid levels for the individuals for which a pair of SNPs co-occurred, and compared the average to that of the individuals for which they did not co-occur (Figure A). For all three pairs, LDL levels increased in the case of co-occurrence. In addition, for *CELSR2*+*LPL* also HDL and TG increased. Furthermore, I checked the relationship between the total number of risk alleles and the lipid levels (Figure B), in individuals with 2, 3 or 4 risk alleles (SNP co-occurence), and with 0 or 1 risk alleles (no SNP co-occurrence). For 2 out of 3 SNP pairs, the LDL levels increased proportionally to the number of risk alleles, whereas for 1 pair (*APOB*+*HMGCR*) LDL increased only in the case of 4 risk alleles. For *CELSR2*+*LPL*, also the TG increased proportionally to the risk alleles.

I then performed the same analysis for those interactors of my co-RNAi screen, which are mapped close to SNPs that were present on the Exome Chip, or had proxies that were present on it (*PCSK9*+*TMEM57*, *LDLR*+*MLXIPL*, *LDLR*+*HAVCR1*, *MLXIPL*+*HAVCR1*), as well as for the triplet *LDLR*+*MLXIPL*+*HAVCR1*. For the three pairs *LDLR*+*MLXIPL*, *LDLR*+*HAVCR1*, *MLXIPL*+*HAVCR1*, as well as for the triplet *LDLR*+*MLXIPL*+*HAVCR1*, the LDL levels increased proportionally to the number of risk alleles. Specifically for *LDLR*+*HAVCR1*, LDL levels rised from negative residual values to zero. In the case of *MLXIPL*+*HAVCR1*, this was also true for TG levels. Interestingly, for *PCSK9*+*TMEM57* the LDL levels significantly increased when one allele of each SNP was present, but decreased when there was homozygosity for at least one of two SNPs (3 or 4 risk alleles) (see Fig. 36).

Figure 36. SNP co-occurrence effect on lipid levels for the 4 gene pairs and 1 triplet from the gene interactions of my co-RNAi screen that are mapped close to SNPs on the Exome Chip used for the Bioimage cohort

(A) The lipid levels (HDL, LDL, TG) in individuals where both SNPs are present (2 SNPs) (or 3 SNPs for LDLR_MLXIPL_HAVCR1) versus individuals carrying 1 of the two SNPs (1 SNP) or no SNP, for the four gene pairs and one triplet that were interactors in my co-RNAi screen and are mapped close to SNPs of the Exome Chip used for genotyping in the Bioimage study. (B) The lipid levels (residuals) for HDL, LDL and TG (y axis) are plotted against the total number of LDL-increasing alleles (x axis), in cases of SNP co-occurence (2, 3 or 4 risk alleles)) and non co-occurence (0 or 1 risk alleles), for the same gene pairs. For the triplet LDLR_MLXIPL_HAVCR1, SNP co-occurrence: 3, 4, 5 or 6 risk alleles, while non co-occurrence: 2 risk alleles. The gene mapped closest to each SNP is shown above the SNP ID.


A
PCSK9_TMEM57
exm62588_exm.rs12027135
Lipid residuals
Total SNPs
no SNPs
1 SNP
2 SNPs
HDL
LDL
TG
HAVCR1_LDLR
exm.rs6882076_exm.rs6511720
no SNP
HAVCR1_MLXIPL
exm.rs6882076_exm625685
LDLR_MLXIPL
exm.rs6511720_exm625685
LDLR_MLXIPL_HAVCR1
exm.rs6511720_exm625685_exm.rs6882076
3 SNPs
B
Total risk alleles

4 DISCUSSION

In the present study, a combinatorial RNAi screening approach was used to explore gene interaction networks underlying human cholesterol regulation, with respect to CVD. The identification through GWA studies of genes associated with lipids and CVD has only to a minor extent explained the genetic architecture of common disorders such as dyslipidemias and CVD. A probable reason for this is that common disorders are assumed to arise from a complex interplay between many genes (Farhan and Hegele 2013); therefore genes need to be studied in combination, rather than in isolation. Considering that a deregulation in cholesterol homeostasis, is a well established major risk factor for CVD (Brown and Goldstein 1974, 1986), I envisaged that by studying the genetic networks behind cholesterol regulation I could shed more light on the genetic architecture of CVD.

With a view to investigating the underlying network connectivity of genes associated with the disease, I tested for interactions those genes that had been associated in GWA studies with lipid traits, CAD and/ or MI. Additionally, the genes that were selected to be tested were already shown to have a functional effect on cholesterol regulation (Blattmann et al. 2013). Interestingly however, most of the effector genes identified by Blattmann et al. only showed moderate effects on LDL uptake upon their depletion via RNAi. This finding further supports the notion that the combined action of two -or more- genes, rather than that of individual genes could account for the regulation of the complex cholesterol homeostasis (Pirucello and Kathiresan 2010).

With an aim to enrich the query set with genes carrying low-frequency and rare variants, that are poorly captured by GWA studies, the gene selection was made from a juxtaposition of the functional profiling data (Blattmann et al. 2013) with genotyping data from a recent Exome Chip of 75k individuals (Peloso et al. 2014), which were tested for associations with blood lipid traits.

With respect to reducing the gene combinations, I followed a focused screening design with a set of 30 genes, which were tested for pairwise interactions. Solid-phase reverse transfection in siRNA-spotted microarrays was employed together with a cell-based functional assay to test the effect of combinatorial gene depletion

on the uptake of LDL-cholesterol from cells. Using high-content automated microscopy and computational analysis, 35 gene-gene interactions were identified (at FDR of 1%) out of 435 pairwise gene combinations (excluding reciprocals).

Following up the identified gene interactions, a number of secondary experiments were performed. First, correlation of the gene knockdown level with the functional effect was investigated. Secondly, the effect of the double knockdowns of interacting gene pairs on *LDLR* mRNA and protein levels, as well as on *SREBF1a* and *SREBF2* mRNA levels was assessed. Moreover, a gene-gene interaction model network was constructed based on results obtained, by overlapping them with available information from STRING database.

The sub-network of the three genes; *HAVCR1-LDLR-MLXIPL* was prioritized for more detailed studies. Preliminary results with these candidates suggest a previously not described function for *HAVCR1* and *MLXIPL* in LDL uptake and a collaborative role of these two genes with LDLR in cholesterol homeostasis.

In parallel, 19 variants of a prominent interactor identified by the co-RNAi screen, *HAVCR1* were tested for their effect on LDL uptake upon overexpression of their GFP-tagged constructs. 15 of those variants were rare (MAF<5%). This was done with two aims; (1) to show the significance of rare variants in complex traits and (2) to further evidence a key role of *HAVCR1* in cholesterol homeostasis, suggested from the findings of the co-RNAi screen.

Finally, in an effort to place the identified genetic interactions in a disease context, a cohort of >4,000 individuals in high risk for CVD (Bioimage study (Muntendam et al. 2010)) was tested for SNP-SNP interactions, among the lead SNPs of the genes investigated with the co-RNAi screen. This was performed in collaboration with Heiko Runz (Merck Research Laboratories, US) and the analysis was done by Gulum Kosova.

In this section discussed in detail are: the model system (4.1) and the functional assays (4.2) used in the present study, as well as the candidate gene selection strategy (4.3). Moreover, the screening approach (4.4) and the gene interaction analysis (4.5) are discussed. In addition, the results of the co-RNAi screen are evaluated (4.6), and hypotheses for mechanistic interpretation of the selected sub-network *LDLR-MLXIPL-HAVCR1* are provided (4.7). Next, the effect of overexpression

of *HAVCR1* mutations is discussed (4.8), and finally, the co-occurrence of SNPs of the tested candidate genes is examined (4.9).

4.1 THE USE OF HELA CELL LINE AS A MODEL FOR CHOLESTEROL HOMEOSTASIS

Using a cell line to model human disease can be considered as reductionism. An *in vitro* disease model system allows studying the tissue of interest in isolation from the rest of the body, thus removing effects resulting from the organic, systemic or organismal level. Therefore, the effect on cholesterol regulation that is observed upon the depletion of a gene product in a cell line might not be the same in an animal model, where the cholesterol regulating system of the whole organism is present. Furthermore, the identification of genes that affect cholesterol homeostasis *in vitro* does not necessarily guarantee that these genes will explain the disease risk in humans. However, the co-RNAi screen that was performed in this study was different from other *in vitro* RNAi screens, because it had the advantage of testing genes, which had already been associated with the disease in GWA and genotyping studies. For this reason, the chances that genetic interactions identified with this screen will be causally related to the disease, are much higher than for other completely unbiased RNAi screens. Furthermore, the study of Blattmann et al. (Blattmann et al. 2013) has already demonstrated the potential of the *in vitro* approach to identify cholesterol-regulating genes within GWAS loci.

In this study, HeLa Kyoto cells were used, an epithelial cell line which derived from cervical cancer. HeLa cells have been widely and successfully used in research of many diseases unrelated to cancer, such as AIDS and poliomyelitis etc. These cells have many characteristics that make them the cell line of choice; they are easy to culture, their doubling time is less than that of other cancer cells, they grow in monolayers and they are adherent. HeLa Kyoto is a strongly adherent version of this cell line. These features make them very suitable for high-throughput microscopy. Furthermore, these cells are very well characterized, as they have been recently sequenced (Landry et al. 2013).

Nonetheless, the suitability of HeLa cells for the study of cholesterol regulation, instead of a liver cell line, can be viewed as controversial by some researchers. Importantly however, it has been demonstrated that although the liver plays a

central role in cholesterol metabolism, the extrahepatic tissues are responsible for more than 80% of the total sterol synthesis (Dietschy et al. 1993). Moreover, extrahepatic tissues are responsible for the clearance of about 30% of the LDL cholesterol. For these reasons, HeLa cells are considered to be a satisfactory model system for cholesterol homeostasis modeling. What is more, LDLR expression levels in Hela Kyoto cells are comparable to other liver-derived cell lines, such as Huh7 and HepG2 (not shown here), which are generally used to study cholesterol homeostasis. In addition, most hepatic cell lines are not suitable for automated high-throughput microscopy experiments, due to reasons such as aggregate formation and variable cell morphology. Besides, non-hepatic cell lines (such as fibroblasts, aortic endothelial, or alveolar epithelial cells) have been extensively used in cholesterol studies and our current knowledge of cellular cholesterol trafficking derives mainly from such studies (Ikonen 2008).

One possible downside of the use of HeLa cells for the co-RNAi screen is that out of the 30 genes tested, seven genes were not detected in the transciptome, six out of those were neither detected in the proteome. However, five of the respective proteins (HAVCR1, NCAN, SEZ6L, CXCL12, MYBPHL) were not detected in neither of ten other cell lines (including the liver-derived cell line HepG2) tested in a comparative proteomic analysis (Geiger et al. 2012). An off-target effect for these genes is ruled out, since more than one siRNA targeting each of those showed an effect in Blattmann et al. (Blattmann et al. 2013). Hence, it is likely that these five genes are expressed under detection levels in tissue culture cells.

4.2 ASSESSING FUNCTIONAL ASSAYS

In order to identify genetic interactions, a phenotypic trait has to be evaluated. Binary phenotypic readouts were assessed in early genetic interaction studies, such as cellular fitness. However, analysis of quantitative phenotypes is much more informative, as it allows the detection of more subtle interactions and furthermore their classification into positive and negative ones.

In the present study, the uptake of fluorescently labeled LDL (DiI-LDL) from cells was used as phenotypic readout, which can be quantified. The phenotypic readout was obtained through performance of the LDL-uptake assay, a well established cellular

assay (Gilbert et al. 2009; Pitas et al. 1981), that has been previously applied in high-throughput screens by our group (Bartz et al. 2009; Blattmann et al. 2013). Thus, the LDL-uptake assay was suitable for high-throughput screening as well as for systematic functional analysis of perturbed human cells.

The aforementioned assay measures the ability of cells to endocytose LDL cholesterol, thus clearing it from the bloodstream. Therefore, LDL uptake is a key process to achieve balanced cholesterol levels in the circulation, and subsequently avoid atherosclerosis and CAD. Hence, the LDL-uptake assay closely reflects the role of the cells in regulating extracellular lipid levels and this is why it was selected in the present study.

4.3 EVALUATION OF CANDIDATE GENE SELECTION STRATEGY

The approach followed for candidate gene selection for the co-RNAi screen turned out to be very efficient resulting in many hits. The gene set was carefully selected to contain both known cholesterol regulators as well as genes not previously functionally related to cholesterol homeostasis, and completely uncharacterized genes. Specifically, of the 30 candidate genes, 8 were already linked to cholesterol regulation (*LDLR*, *LDLRAP1*, *PCSK9*, *MYLIP*, *HMGCR*, *LPL*, *APOB*, *APOE*), and 2 additional (*MLXIPL*, *PAFAH1B1*) were linked to lipid homeostasis by GO (Gene Ontology) annotation. The remaining 20 genes had not been previously linked to lipid homeostasis. Moreover, for 5 of those (*CELSR2*, *FAM174A*, *MYBPHL*, *SIK3*, *TMEM57*) no functional information was available in the literature (see Tbl. S5).

Interestingly, genes with a well-defined cholesterol regulatory function, such as *SREBP*, *SCAP*, *INSIG* and *LXR* were entirely absent from GWAS, and therefore from the co-RNAi screen as well. The lack of GWAS-identified disease-associated SNPs in these genes might be either due to the highly detrimental effect of those SNPs, which are therefore removed by purifying selection, or due to their low frequency that renders them unlikely to be detected by GWAS. On the other hand, genes regulating cholesterol removal (such as *ABCA1*, *ABCG5*, *ABCG8*) or storage (such as *LCAT*) were reported in GWAS but did not show a significant effect on LDL uptake or cholesterol levels upon depletion (Blattmann et al. 2013), therefore were not included in the co-RNAi screen.

It has commonly been assumed that strong epistatic interactions are detected between genes with small main effects (Phillips 2008), although contrasting theories support that interactions are more likely to occur between alleles with more severe effects (Crow 1990). With this consideration in mind, I selected for the interaction screen both genes with a small individual effect (such as CELSR2) and genes with a stronger effect (such as LDLR) on LDL uptake and/or total cholesterol levels (Blattmann et al. 2013).

Nonetheless, not all genes with a single phenotypic effect in Blattmann et al. were selected for the co-RNAi screen. Ideally, one could select all the 55 genes that were effectors in the aforementioned study. This would though result in 3025 pairwise combinations to be tested, as compared to 900 that correspond to 30 candidate genes. However, it needs to be pointed out that the current gene selection was not meant to be exhaustive for the gene interactions that affect cholesterol regulation, but rather highly enriched in gene interactions that could be then followed up by more experiments.

Moreover, it should be noted that the candidate genes from Blattmann et al. originate from GWA studies, which are only powered to detect common variants. With an aim to more efficiently address the missing heritability problem (see section 1.1.3.2.1), genes with rare variants associated to lipid traits were also included in the screen. Therefore, the results of the Blattmann et al. were juxtaposed to those of a genotyping study for lipid traits (Peloso et al. 2014). In this study, a cohort of >70,000 individuals was genotyped for LDL-c, HDL-c and TG (Peloso et al. 2014). Specifically, the genotyping was performed with the Illumina HumanExome array, which was built to contain coding sequence variants discovered from the exome sequencing of ~12,000 individuals. In this way, low-frequency and rare variants were associated with lipid traits.

All 30 candidate genes selected for the interaction screen contain or are located adjacent to SNPs identified by GWAS to be associated with blood lipid levels (LDL, HDL, TG, TC), CAD and/ or MI (Blattmann et al. 2013). Namely, these genes are residing in loci associated with cardiovascular disease risk, either directly or indirectly -through association with causal risk factors for the disease. As a

consequence, interactions detected among these genes would be immediately associated with cardiovascular disease risk.

What is more, 23 of these genes (see Tbl. S2) contain rare variants which showed a significant association with lipid traits in the aforedescribed genotyping study (Peloso et al. 2014). The analyzed cohort consisted at large of participants of cardiovascular disease studies, in which affected individuals presented with atherosclerosis, MI and/or CAD. Hence, the genes investigated in the co-RNAi screen performed in the present study have a direct relation with CVD.

4.4 EVALUATION OF THE CO-RNAI APPROACH

The concept of studying gene-gene interactions via the systematic co-depletion of genes and the analysis of a quantitative phenotype has already been implemented on metazoan cells, including *C. elegans* (Lehner et al. 2006; Tischler et al. 2006), *Drosophila* cells (Billmann et al. 2016; Horn et al. 2011), mouse (Roguev et al. 2013) and human cell lines (Barr and Bakal 2015; Laufer et al. 2013; Wang et al. 2014). Using co-RNAi, multiple genes can be targeted for depletion at the same time, thus enabling the observation of phenotypic outcomes that arise from the combined contribution of functionally connected genes. This is a powerful approach to reveal functional relationships between previously unrelated genes, as well as to yield important information concerning the biological pathways in which the genes take part. Moreover, by constructing the genetic interaction profile of a gene, that is the total of genetic interactions in which it participates, the function of a previously uncharacterized gene can be inferred.

One aspect of the co-RNAi approach that needs to be taken into account is that we frequently obtain only partial depletion, rather than complete abolishment of the gene product through RNAi-mediated knockdowns. The inhibition of gene expression induced by RNAi depends on the knockdown efficiency of the siRNA used, the stability of the gene product, as well as regulation of gene expression by feedback mechanisms. Intermediate gene expression levels lead to the emersion of partial phenotypes (hypomorphs) that would not be observed otherwise. This has both drawbacks and benefits; one drawback is that the outcome of the simultaneous complete depletion of the two gene products cannot be observed. This would show

the effect of the complete absence of function of both genes. On the other hand, however, the incomplete knockdowns result in substantially reduced gene expression levels but sufficient for their viability, which is crucial in the case of essential genes. In addition, partial phenotypes have been shown to frequently yield biologically meaningful genetic interactions (Schuldiner et al. 2005). Moreover, the intermediate expression levels achieved through RNAi result in phenotypes very closely resembling those of patients carrying partial loss-of-function mutations in the respective genes. Thus, the RNAi-mediated reduction of gene expression, which can be tuned, can lead to different levels of residual gene activity, which is physiologically relevant and therefore its effects can be very informative.

4.5 EVALUATION OF THE GENETIC INTERACTION ANALYSIS

The statistical analysis of genetic interactions is challenging in the aspects of their definition, detection and interpretation. A genetic interaction is considered as an unpredictable phenotype that emerges from the combination of two or more alleles. However, different mathematical definitions have been provided, as it is not straightforward to predict in which way individual alleles combine to produce a phenotype (Mani et al. 2008).

In this study, one of the most commonly used approaches was used, deviation from additivity. This approach is based on the original definition of statistical epistasis by Fisher in 1918, in which the effect on a quantitative phenotype of two alleles at different loci deviates from additivity in a statistical model (Cordell 2002; Fisher 1918). Under an additive model, each allele is assumed to have an additive contribution to the phenotype, in such a way that the phenotype of the double mutant equals the sum of the two mutants, in the case of no genetic interaction. In this context, in order to detect genetic interactions in the present study, the observed double knockdown phenotype was compared to the expected one, which would equal the sum of the two single ones.

Different statistical models and computational methods have been employed for the detection of genetic interactions. For the modeling of quantitative traits, a popular choice is linear regression (Gilbert-Diamond and Moore 2011). Here, I applied linear regression to model the ability of cells to endocytose cholesterol, which is a

quantitative phenotype. In linear regression analysis, the relationship between a continuous dependent ("*outcome*") variable (Y) and one (simple linear regression) or more (multiple linear regression) continuous independent ("*predictor*") variables (X_1, X_2, etc.) is modeled. Here, multiple linear regression was used to model the amount of fluorescently-labeled cholesterol uptake (Y), as a function of two single knockdown phenotypes (X_1, X_2). In this model, which corresponds to an additive model, the expected outcome phenotype for a pair of non-interacting genes is the sum of the two single-gene effects (main effects).

In the current combinatorial RNAi screen two different approaches were followed to estimate the main effects of single siRNAs. In the first one, the main effect of a query siRNA was estimated from the combinations of the siRNA with the nontargeting control partner siRNA (NEG9), whereas in the second approach the main effects were estimated from all double RNAi measurements that included the query siRNA. The first approach is used when among the investigated gene siRNAs there are many with measurable knockdown effects, whereas the second one is used when the majority of them show negligible effects (Horn et al. 2011; Laufer et al. 2014). Since most of the siRNAs did not have a strong phenotypic effect, both approaches were applicable in this case.

Genetic interactions can be easily modeled with linear regression, using a standard desktop computer and a statistical package, such as the package *"MASS"* of the R software environment, which was used here. Moreover, the mathematical theory behind linear regression analysis of interactions is very well characterized.

Once a model for genetic interactions has been established, the fitted model, as well as the detected interactions, need to be interpreted. That is to say, etiological inferences need to be made from the computational model, in order to deduce what is the biological relevance of the interactions detected. More specifically, it should be clarified how two genes combine to associate with a trait or a disease. At this point, it is critical to stress that the genetic interactions tested for with this analysis are statistical interactions and not physical interactions (Cordell 2009; Phillips 1998, 2008). In other words, the fact that two genes have an unexpected combinatorial effect on an observed phenotype does not necessarily mean that the corresponding

gene products physically interact. The exact nature of the interactions has to be further investigated.

4.6 RESULTS OF THE CO-RNAI SCREEN

The simultaneous co-depletion of two different gene products with RNAi showed both expected as well as unexpected results. For instance, the majority of siRNAs had a down-regulating effect on LDL uptake when co-transfected with *LDLR* siRNA, which was similar to the effect of *LDLR* single knockdown, demonstrating that the effect of *LDLR* was dominant over the other gene (see Fig. 14). This result is in accordance with the essential role of *LDLR* on LDL-uptake, since in absence of functional LDL receptors at the plasma membrane, LDL cannot be efficiently endocytosed (Goldstein et al. 1985). Only in two cases the co-knockdown of *LDLR* with a second gene did not result in a decreased LDL uptake; that of *HMGCR* and *PVRL2*.

HMGCR (3-hydroxy-3-methylglutaryl-CoA reductase) encodes for the rate-limiting enzyme in cholesterol biosynthesis pathway. Therefore, a knockdown of *HMGCR* would decrease cellular cholesterol synthesis and promote the uptake of cholesterol from the cell to compensate for low cellular cholesterol levels. This is in agreement with the results of my screen, where knockdown of *HMGCR* together with the negative control siRNA, or in combination with most other genes upregulated LDL uptake. The upregulating effect of *HMGCR* knockdown is counteracted by the downregulating effect of *LDLR* knockdown, and this results to a neutral effect, as expected.

PVRL2 (Poliovirus receptor-Related protein 2), on the other hand, encodes for a membrane glycoprotein, which serves as a plasma membrane component of adherence junctions and has no described connection to cholesterol homeostasis. Its single knockdown had no significant effect on LDL uptake, but neither did however the co-knockdown with *LDLR*. This gene pair was identified as a positive interaction, as it alleviated the inhibitory effect of *LDLR* depletion on LDL uptake. As the function of this gene is at large uncharacterized, I cannot generate hypothesis to explain this unpredictable result.

Unexpectedly, the depletion of *LDLRAP1* resulted to up-regulation of LDL uptake, which is in agreement with the previous study (Blattmann et al. 2013) but in contrast to existing literature on *LDLRAP1*. *LDLRAP1* (alias; *ARH*) encodes for a known adaptor protein for LDLR, which promotes the internalization of the LDLR-LDL complex in hepatocytes, via clustering LDLRs to clathrin-coated pits (Garuti et al. 2005; He et al. 2002; Sirinian et al. 2005). In addition, it has been shown that mutations in this gene lead to autosomal recessive hypercholesterolemia (ARH) (see section 1.3.2.3) due to defective internalization of the receptor-ligand complex (Cohen et al. 2003; Garcia 2001). Therefore, it is expected that *LDLRAP1* depletion would lead to down-regulation of LDL uptake. One possibly explanation for the unexpected result might be that LDLR endocytosis in HeLa cells might be regulated differently. For example, it has been shown that in hepatocytes LDLRs are dispersed on the plasma membrane, whereas in fibroblasts they are clustered in coated pits (He et al. 2002). Moreover, DAB2, the alternative adaptor protein for LDLR, is expression at higher levels in HeLa cells than in hepatocytes (Maurer and Cooper 2006). Therefore, *DAB2* could be substituting *LDLRAP1* in the absence of the latter, and a co-depletion of both LDLRAP1 and DAB2 is probably required to give an effect on LDL uptake in non-hepatic cells (Eden et al. 2007).

The co-depletion of *MYLIP* with most other genes resulted to an up-regulation of LDL uptake, as expected. *MYLIP* (alias; *IDOL*: Inducible Degrader of the LDLR) encodes for an E3 ubiquitin ligase that promotes LDLR degradation, via its ubiquitination (Zelcer 2009; Zhang et al. 2011). Therefore, depletion of MYLIP would lead to decrease LDLR degradation, therefore increased LDLR concentration at the plasma membrane and consequently increased LDL uptake. The up-regulating effect is reversed upon co-depletion with *LDLR*, as expected, showing the dominance of *LDLR* effect over *MYLIP*.

Unexpectedly, the knockdown of *PCSK9* does not have an upregulating effect on LDL uptake, and its combinatorial knockdown with the other 29 genes has both upregulating and downregulating effect. *PCSK9* (Proprotein Convertase Subtilisin/Kexin type 9) encodes for a protease which promotes the degradation of LDLR in lysosomes, after its internalization (Zhang et al. 2007). Consequently, depletion of *PCSK9* would be expected to result in decreased LDLR degradation, and

subsequently increased LDL uptake. A possible reason for this contradiction to the literature results might be the use of a non-hepatic cell-line. PCSK9 has been shown to promote LDLR degradation in hepatocytes, however the role of the gene in our model system had not been studied so far.

4.6.1 IDENTIFIED GENE-GENE INTERACTIONS

The reliability of the screening approach and the appropriateness of the statistical model applied for the analysis of the co-RNAi screen data were corroborated by the identification of a well-known interaction pair; *LDLR+LDLRAP1*, which indicates that biologically meaningful interactions can be derived with this approach. The interaction between LDLR and LDLRAP1 (LDL Receptor Adaptor Protein) represents a well established genetic and physical interaction, as LDLRAP1 is known to interact with LDLR to achieve efficient endocytosis of LDLR (Garcia 2001; He et al. 2002) (see section 1.3.2.3).

The primary analysis approach (control-based) of the interaction screen revealed 35 interactions among 24 (*APOB*, *APOE*, *BCAM*, *CELSR2*, *CXCL12*, *HAVCR1*, *HMGCR*, *LDLR*, *LDLRAP1*, *LPL*, *MLXIPL*, *MYBPHL*, *MYLIP*, *NCAN*, *PAFAH1B1*, *PCSK9*, *PVRL2*, *SEZ6L*, *SIK3*, *SORT1*, *TMEM57*, *TOMM40* and *ZNF259*) of the 30 genes tested, at p-value<10^{-2}. Seven of the candidate genes tested (*BAZ1B*, *BCL7B*, *CBLC*, *FAM174A*, *PAFAH1B2*, *TM6SF2* and *WDR12*) did not participate in any interaction. Among the 24 "interactors" (genes participating in interactions), were all 9 genes that had been previously linked to cholesterol regulation (see Table S5). The remaining 15 genes had no previously described cholesterol-regulatory function, except for the Blattmann et al. study, which showed ane effect of their depletion on LDL uptake and/or cholesterol levels (Blattmann et al. 2013).

Of the 35 identified interactions, approximately one-third (11) resulted in an upregulation of LDL uptake (positive interaction effect), and two-thirds (24) downregulated LDL uptake (negative interaction effect) (see Tbl. 4). This finding is interesting, considering that the candidate gene set was equally enriched in genes with an individual positive (8 genes) or negative (8 genes) effect on LDL-uptake assay in Blattmann et al. (see Table S1). The prevalence of negative interactions among the screen hits is easily explained, taking into account the fact that these genes derive

from GWA studies and are associated to disease either directly (CAD/MI) or indirectly (lipid traits). It is therefore expected that depletion of the function of these genes would lead to decreased uptake of LDL-cholesterol from the cells, which results in increased circulating LDL levels, a risk factor for CAD/MI.

The alternative analysis approach (sample-based) revealed significantly fewer interactions; only 2 gene-gene interactions were identified at the same threshold with the primary analysis (p-value<10^{-2}), and 7 at a lower threshold (p-value<$5x10^{-2}$). There was a 100% overlap between the hits of the two methods, at the same thresholds (see 3.1.7). However, the *LDLR+LDLRAP1* interaction, as well as many other interactions that were later confirmed with secondary experiments were missed out with this analysis. Therefore, the first analysis approach was considered as more appropriate for the existing dataset.

4.6.2 GENE INTERACTION VALIDATION

In order to correlate the phenotypic effects with the knockdown efficiency, as well as to assess the effect of the detected interactions on the expression levels of *LDLR* and *SREBP*, the 32 double knockdowns that were identified as gene-gene interactions, along with the respective single knockdowns, were repeated with forward, liquid-phase transfection, in a multiwell-plate format that allowed for these experiments. Surprisingly, for many genes (13 out of 23 taking part in interactions) more penetrant gene knockdowns were observed than in the screen. As in many cases the single knockdown phenotypes were stronger upon forward transfection, the double knockdowns were not able to further potentiate the effect of the singles, and therefore the interaction could not be detected (see Fig. S1a).

For this reason, the siRNA concentrations had to be adjusted for those genes showing a more penetrant knockdown in forward transfection, in order to achieve the same single knockdown phenotypes as in the screen, and consequently replicate the initially detected interactions (see Fig. S1b).

There might be various reasons why siRNA-mediated knockdowns were shown to be more penetrant in forward, liquid-phase than in reverse, solid-phase transfection. First, a key difference between forward and reverse transfection is the condition of cells at the time of transfection. In solid phase transfection, the cells are transfected

at the moment that they are plated, as opposed to liquid-phase transfection, whereby the cells are transfected one day after they are plated. Consequently, in the two approaches the cells are in different condition at the time of the transfection. In addition, in reverse transfection the cells share the same growth medium in a spotted array, therefore secreted signals from cells on one spot might affect the phenotype of cells at a different spot.

After adjusting the siRNA concentrations, 20 out of 35 gene interactions were reproduced with liquid-phase transfection (see Tbl. 4 and Fig. 17). These 20 hits had a statistically significant interaction effect, which had the same direction as in the screen (up- or down-regulating of LDL uptake). Of the 20 replicated interactions, nearly one-third (6) had a positive effect on LDL update (positive robust Z score), while nearly two-thirds (14) had a negative effect on LDL update (negative robust Z score). Of those 20 interactions, one-quarter (5) were positive (positive interaction effect), while three-quarters (15) were negative (negative interaction effect) (see Tbl. 4 and Fig. 17). In addition, one more interaction that did not initially score as statistically significant, but seemed interesting to us (see section 3.3) was independently confirmed (*LDLR+HAVCR1*). Among the replicated gene interactions, the prevalence of negative interactions was obvious. As was aforementioned, this finding is rather expected due to the association of the tested genes with disease, a risk factor for which is the decreased clearance of circulating LDL by the cells.

4.6.3 Correlation of Knockdown Efficiency with Phenotypic effect

The model system that was utilized here for the study of genetic interactions is very complex and dynamic, as evidenced by several observations. First, the effect of many siRNAs that demonstrated higher knockdown efficiency (such as *BCAM*, *PAFAH1B1*, *SIK3* in Tbl. 5) was milder than that of siRNAs resulting to lower knockdown (such as MLXIPL, *TMEM57* and *HMGCR* in Tbl. 5). Taking into account that all siRNAs were transfected for 48 hours, different kinetics of gene expression and transcript processing might be causing this effect. Namely, the products of the genes tested could have dissimilar half-lives, and the period required for the attainment of new mRNA levels after siRNA-mediated degradation is related to those. Within a high-throughput RNAi screen it is not feasible to achieve knockdowns at different time

points, because the functional assay is performed at the same time for a microarray spotted with many different siRNAs. However, for some genes longer knockdown periods might be required to achieve product depletion, which exceed that preceding the phenotypic readout. Therefore, measurements at different intervals might be needed in order to properly evaluate the efficiency of the siRNAs.

Secondly, it was noticed that reduced amount of siRNA resulted in substantial increase/decrease of the effect for some genes (see Fig. S1b), while for others the phenotype remain at large unaffected. Moreover, the phenotypic effect of some siRNAs differed between the screen and the liquid-phase transfections (e.g. *CXCL12*+*PAFAH1B1* in Tbl. 4 and Tbl. 5). These findings demonstrate that the cholesterol-regulating model system is sensitive and adapts to perturbations in order to achieve balance.

The latter can be understood if we consider that multiple intertwined feedback mechanisms control cholesterol homeostasis in the cell. An outstanding example of feedback process is the aforedescribed *INSIG-SCAP-SREBP* system (see 1.3.2.2). Decreased cholesterol levels results in activation of the system, which turns on the cholesterol biosynthetic pathway (Brown and Goldstein 1997). The same system increases the expression of *LDLR*, leading to increased cholesterol uptake from the cell. Once enough cholesterol has been synthesized, it functions to shut down the *INSIG-SCAP-SREBP* system, thus closing the feedback loop (Brown and Goldstein 2009). This paradigm demonstrates the potential of the cellular cholesterol-regulatory machinery to efficiently respond to external stimuli through complicated processes.

4.6.4 Double Knockdowns demonstrate the irreplaceable role of LDLR

Importantly, it was observed that in most cases the combinatorial depletion of *LDLR* with a second gene could not further potentiate the down-regulating effect of *LDLR* single knockdown on either LDL uptake or LDLR expression levels. As was already mentioned (section 4.6), the majority of siRNAs had a similar effect on LDL uptake to that of *LDLR* single knockdown when co-transfected with *LDLR* siRNA (see Fig. 14). This was also true for two of the four interactions of *LDLR* that were identified by the co-RNA screen: *LDLR*+*LDLRAP1* (Fig. S2.7) and *LDLR*+*MLXIPL* (Fig. S2.10). Only in two

cases LDL-uptake was more inhibited upon double knockdown: *LDLR*+*NCAN* (Fig. S2.1) and *LDLR*+*HAVCR1* (Fig. S2.21). However, the co-knockdown of *LDLR* with *HAVCR1* could not further decrease *LDLR* mRNA (Figs. 21 and 22) or protein levels (Figs. 23 and 24). Neither could the co-knockdown with *LDLRAP1* (see Figs. 21, 23, 24). Nonetheless, co-depletion of *LDLR* with *NCAN* or *MLXIPL* resulted in lower LDLR transcript levels (Fig. 22).

These findings demonstrate that the effect of *LDLR* on LDL uptake is dominant over most of the other genes tested. Upon *LDLR* depletion by siRNA, the amount of LDL receptors on the plasma membrane exceeds the lower threshold, and cannot be decreased more by co-knockdown with an siRNA targeting a second gene. This rationale is in accordance with the essential role of *LDLR* on LDL-uptake, since in absence of functional LDLRs at the plasma membrane, LDL cannot be efficiently endocytosed (Goldstein et al. 1985). Furthermore, the results from LDLR protein expression experiments again illustrate the epistatic effect of *LDLR* itself on its post-transcriptional regulation. Hence, the unexpected effect of *LDLR* co-depletion with *NCAN* or *MLXIPL* on the transcriptional regulation of *LDLR* (Fig. 22) is considered very interesting and are discussed later (see sections 4.6.6 and 4.6.7).

4.6.5 CONSTRUCTION OF A MODEL GENE INTERACTION NETWORK

Aiming to uncover potential functional relationships among the identified interactors of the co-RNAi screen, a model gene-gene interaction network was constructed among all 30 genes tested in the screen, by overlapping the 21 validated gene interactions with known interactions from the STRING database. The overlap of the validated *LDLR*+*LDLRAP1* gene interaction with an experimentally determined interaction from STRING showed that the screening approach employed can indeed identify interactions between cholesterol regulating genes (see section 4.6.1). The fact that only one out of eight experimentally determined interactions (pink lines in Fig. 29) from STRING was identified with the co-RNAi screen, could be due to the use of a single phenotypic readout for the identification of gene-gene interactions. Namely, only the effect on LDL uptake of combinatorial gene depletion was assessed, which means that any interactions implicated in cholesterol regulation but not implicitly in the endocytosis of LDL, would be probably missed out. Furthermore,

the interactions shown in STRING are not related to cholesterol metabolism, but to different cellular functions or processes. That is to say that the reported interactions were mapped under completely different experimental conditions and phenotypic readouts, and most likely concern other aspects of the cellular function. Moreover, false results from earlier work should also not be excluded. Therefore, an overlap is not necessarily expected.

Noteworthy, in the constructed genetic interaction network most genes interact with only one or two genes, whereas a small number of genes have many interactions, such as *HAVCR1*, which interacted with 6 other genes, or LDLR and *LDLRAP1* with 4 other genes each. This finding is consistent with the non-uniform topology of most cellular networks, where most of the nodes are sparsely connected having one or two links, whereas a few nodes are highly connected, having a very large number of links, and therefore serve as network "hubs" (Barabasi and Oltvai 2004). Thus, the genes *HAVCR1*, *LDLR* and *LDLRAP1* can be considered as "hubs" in the present network, which probably means that their role in the cholesterol-regulatory pathways is quite central. Although the majority of genetically interacting genes belong to the same pathway or biological process, network "hubs" usually interact with many functionally diverse genes and tend to be pleiotropic (Costanzo et al. 2010). This means that each of these three aforementioned "hub"- genes is likely to take part in a number of distinct pathways/ processes affecting LDL endocytosis.

Noteworthy, the genetic interaction profile of a gene, namely the total of genetic interactions in which a gene participates, can give important information for the function of the gene, since genes taking part in the same pathway usually have similar interaction profiles (Costanzo et al. 2011). For example, I could hypothesize that *HAVCR1* might take part in the same pathway as *LDLRAP1*, since their interaction profiles are similar -they both interact with *BCAM*, *LDLR* and *SORT1*, as well as with each other.

4.6.6 HYPOTHESES FOR MECHANISMS OF GENETIC INTERACTIONS

With regards to building hypotheses on the mechanisms of the identified gene interactions, a number of secondary experiments were performed. Asking if the

interacting genes exert their effects via modulation of *LDLR* transcription and/or translation, the effect of the 21 validated interactions on *LDLR* mRNA and protein levels was measured. In addition, to determine whether the gene interactions regulate cholesterol homeostasis through the INSIG-SCAP-SREBP system via affecting *SREBF* transcription, the effect of the 21 gene-gene interactions on *SREBF* mRNA was also tested.

The INSIG-SCAP-SREBP system, which regulates cholesterol homeostasis at a transcriptional level, has already been described (see section 1.3.2.2). The *SREBF* family in human consists of *SREBF*-1a, *SREBF*-1c and *SREBF2*, of which *SREBF*-1a and *SREBF2* upregulate cholesterol synthesis (Goldstein and Brown 2015). Therefore, the effect on both *SREBF1* and *SREBF2* mRNA levels was measured.

In the following part a few interesting interactions are discussed in detail. First I discuss three interactions (*LDLR*+*LDLRAP1*, *LDLR*+*NCAN* and *SORT1*+LDLRAP1) of the "hub" genes of the identified network; *LDLR* and *LDLRAP1*.

The pivotal role of *LDLR* and the accessory role of *LDLRAP1* in cholesterol uptake were already described (section 1.3.2.3). Co-depletion of *LDLR* with *LDLRAP1* resulted in reduced LDL-uptake, at levels similar to those of single *LDLR* knockdown (Fig. S2.7). Hence, the interaction identified between the two genes is epistatic, and was reflected also at the post-transcriptional level of *LDLR* expression (Fig. 24). Moreover, the double knockdown increased *SREBF2* mRNA levels (Fig. 25 and 27). This finding is expected taking into account that decreased endocytosis of LDL caused by the co-depletion of *LDLR* and *LDLRAP1* results in reduced cellular cholesterol levels, which would need to be compensated for. An activation of the SCAP/SREBP system would increase *LDLR* expression as well as cholesterol synthesis, thus restoring LDL endocytosis and replenishing the cholesterol pool (Goldstein et al. 2006). Normally, the SCAP/SREBP pathway could be turned on by simply transporting the available SCAP/SREBP complexes from the ER to the Golgi and cleaving SREBPs. However, upon high demand for cholesterol, an increase in *SREBF2* synthesis would be required, as observed here.

NCAN (Neurocan) is located on 19p13.11 locus, which is associated with Tc levels. The co-depletion of *NCAN* with *LDLR* further potentiated the inhibitory effect of *LDLR* depletion on LDL uptake (Fig. S2.1). Additionally, the combinatorial knockdown of

the two genes further reduced *LDLR* mRNA levels (Fig. 22). Moreover, it alleviated the diminishing effect of the two single knockdowns on *SREBF2* mRNA (Fig. 27). These findings suggest that the interaction of *NCAN* with *LDLR* might down-regulate *LDLR* transcription, via inhibition of *SREBF2* transcription, thus leading to reduced LDL uptake. *NCAN* encodes for a surface proteoglycan, which may be involved in neuronal adhesion and neurite growth during development, via binding to neural cell adhesion molecules (Rauch, Feng, and Zhou 2001). Hence, one could alternatively hypothesize that NCAN depletion might interfere with the adhesion of the cells, thus indirectly affecting various other unrelated processes, including LDL endocytosis.

SORT1 resides at the genomic locus 1p13, which has the strongest association to LDL-cholesterol, and is associated with CVD as well. Here, the co-depletion of *SORT1* and *LDLRAP1* further potentiated the inhibitory effect of SORT1 depletion on LDL uptake (Fig. S2.6). *SORT1* encodes for a transmembrane receptor of many ligands, including lipoprotein lipase (*LPL*), a triglyceride hydrolase (Nielsen et al. 1999). The main function of sortilin is the binding of ligands in the Golgi or at the plasma membrane and their trafficking to the lysosome. Loss-of-function studies in $LDLR^{-}/^{-}$ mice demonstrated that sortilin binds LDL at the cell surface, contributing to the non-LDLR-mediated uptake of LDL, and mediates its lysosomal degradation (Strong et al. 2012).

Taking the above into account, the combined effect of *SORT1* and *LDLRAP1* on LDL endocytosis can be explained, since SORT1 takes part in non-LDLR-mediated uptake of LDL (Patel et al. 2015), whereas LDLRAP1 is involved in LDLR-mediated uptake of LDL (Garuti et al. 2005; Maurer and Cooper 2006), which are two parallel and convergent pathways (see Fig. 9). In addition, the co-knockdown of *SORT1* and *LDLRAP1* resulted in the reduction of both *SREBF1* (Fig. 26) and *SREBF2* mRNA (Fig. 27). This suggests that co-depletion of the two gene products might downregulate LDL uptake by inhibiting LDLR synthesis, via reduction of *SREBF1* and *SREBF2* transcription.

Sortilin has been also implicated in atherosclerosis, though existing reports are quite contradictory (Mortensen et al. 2014; Patel et al. 2015). Hence, *SORT1* appears to have a key role in LDL metabolism and CAD (Tall and Ai 2011), and the clarification of its synergistic contribution with *LDLRAP1* to both of them is needed.

Secondly, two more interesting interactions are now discussed; *PCSK9*+*TMEM57*, *NCAN*+*TOMM40*. *PCSK9* resides at the locus 1p36.11, whereas *TMEM57* at 1p32.3. *PCSK9* (proprotein convertase subtilisin/kexin type 9) encodes for a degrader of LDLR, while it also decreases the recycling of LDLR to the plasma membrane (see section 1.3.2.3). *TMEM57* is a largely uncharacterized gene. The combinatorial knockdown of *PCSK9* with *TMEM57* strongly upregulated LDL uptake, inverting the effects of both single depletions (Fig. S2.20). This finding suggests that *TMEM57* might also affect the recycling of LDLR. Furthermore, the double knockdown of the two gene products reversed the inhibitory effect of both single knockdowns on *SREBF1* transcription (Fig. 26), while it also alleviated the diminishing effect of both singles on *SREBF2* mRNA (Fig. 27). Based on these results, I could hypothesize that the interaction between *PCSK9* and *TMEM57* might increase LDL uptake by affecting SCAP/SREBF-mediated LDLR synthesis.

TOMM40 resides at the locus 19q13.32 and NCAN at the locus 19p13.11. *NCAN* was previously described. *TOMM40* (Translocase Of Outer Mitochondrial Membrane 40) encodes for the channel-forming subunit of the translocase of the mitochondrial outer membrane (TOM) complex, which imports protein precursors into mitochondria. The combined depletion of *NCAN* and *TOMM40* decreased LDL uptake (see Fig. S2.4) as well as LDLR mRNA levels (Fig. 22). In addition, it decreased *SREBF2* mRNA levels, having a similar effect to that of both single knockdowns (Fig. 27). These results suggest that the interaction of the two genes might reduce LDL uptake through down-regulation of LDLR transcription, induced via inhibition of *SREBF2* transcription. Based on the function of *TOMM40*, I could hypothesize that its interaction with *NCAN* might affect the trafficking of LDLR through mitochondria, subsequently disturbing its recycling back to the plasma membrane.

The correlation of LDL-uptake with LDLR mRNA and protein as well as SREBF1 and SREBF2 mRNA levels for all 21 validated gene interactions is shown in Tbl. S18. For those that are not discussed here, the results were complicated and inconclusive. Therefore, further investigations are required for the elucidation of the underlying mechanisms.

4.6.7 THE GENE INTERACTION SUB-NETWORK: *LDLR-MLXIPL-HAVCR1*

A very interesting sub-network of the gene interaction model network, which was selected for follow up, is that of *LDLR-MLXIPL-HAVCR1*. In this network, each gene has a combinatorial effect with the other two on LDL internalization. Specifically, the combinatorial knockdown of *LDLR* with *MLXIPL* (Fig. S2.10) or with *HAVCR1* (Fig. S2.21) is potentiating the negative effect of *LDLR* knockdown on LDL uptake. Conversely, the combinatorial knockdown of *HAVCR1* with *MLXIPL* (Fig. S2.5) is reversing the positive effects of the two single knockdowns, leading to a down-regulation of LDL uptake. In order to construct hypotheses on the mechanistic interpretation of these interactions, one needs to take into account what has already been reported concerning the function of these genes.

LDLR is a well-described gene that encodes for the receptor of LDL, and is responsible for its endocytosis, as was already described (see section 1.3.2.3).

MLXIPL (MLX Interacting Protein Like) resides on locus 7q11.23, which is strongly associated with plasma triglycerides. *MLXIPL*, (also known as ChREBP: Carbohydrate Responsive Element Binding Protein) encodes for a transcription factor that converts excess hepatic carbohydrates to lipids (Cha and Repa 2007; Stoeckman et al. 2004). The mechanism of action of MLXIPL is similar to that of SREBP, the main difference being that MLXIPL is sensitive to glucose, whereas SREBP is sensitive to cholesterol. Specifically, upon increase of glucose levels MLXIPL upregulates the transcription of genes associated with glycolysis, lipogenesis and fatty acid synthesis (Park et al. 2014). Noteworthy, the depletion of *MLXIPL* alone resulted in a 0.25 fold decrease in *SREBF2* mRNA, whereas the combinatorial depletion with *LDLR* resulted in a slight increase of *SREBF2* mRNA (Fig. 27 and 25). Taking the above into account, I hypothesize that *MLXIPL* might function as a activator of *SREBF2* transcription in the presence of *LDLR*, while this effect is abolished upon *LDLR* depletion. Namely, when there is sufficient amount of LDL receptors on the plasma membrane, and subsequently adequate LDL-c is internalized, MLXIPL might function as an “on” switch for the SCAP-SREBP system. On the contrary, upon depletion of LDLR, MLXIPL might block *SREBF2* transcription, thus allowing for the deactivation of SCAP-SREBP.

MLXIPL was shown here to exert a combinatorial effect with *LDLR* on *LDLR* mRNA, further decreasing its levels (Fig. 22). However, *MLXIPL* depletion resulted in

increased LDL uptake (Fig. S2.10), which correlated with higher LDLR protein levels (Fig. 23). To explain these apparently contrasting results one has to take into account that mRNA and protein levels do not necessarily correlate. Various factors, including post-transcriptional modifications and protein half-lives could account for discrepancies between the two. The above findings could lead to the hypothesis that MLXIPL might function as a transcription factor of *LDLR.* Together with the aforedescribed results on *SREBF2* mRNA, I could hypothesize that *MLXIPL* might participate in the regulation of *LDLR* transcription via controlling the *SREBF2* component of the SCAP-SREBF pathway.

The genetic interaction identified by the screen between *LDLR* and *MLXIPL*, together with the secondary findings is very interesting, because *MLXIPL* has been demonstrated to protect against atherosclerosis (Sarrazy et al. 2015). In particular, it was shown that *MLXIPL* deletion in $LDLR^{-}/^{-}$ mice resulted in faster atherosclerosis progression (Sarrazy et al. 2015).

HAVCR1 (Hepatitis A Virus Cellular Receptor) -also known as *TIM1* (T-Cell Immunoglobulin Mucin Receptor 1) or *KIM1* (Kidney Injury Molecule 1)- resides at locus 5q33.3, which is strongly associated with plasma LDL and Tc levels. *HAVCR1* encodes for a receptor for hepatitis A virus and other viruses (s.a. Ebola virus), as well as for T-cells. Interestingly, it has been demonstrated that HAVCR1 can bind and internalize oxidized -and possibly also native- LDL, though the molecular mechanism has not been clarified (Ichimura et al. 2008). Hence, the synergistic effect of *HAVCR1* with *LDLR* on LDL uptake identified by co-RNAi (Fig. S2.21) could lead to the hypothesis that HAVCR1 might act as an alternative receptor for LDL, with lower affinity for LDL than LDLR.

With a view to investigating this hypothesis, I performed some follow-up experiments for *LDLR* and *HAVCR1*. In order to examine if HAVCR1 could function as an alternative receptor for LDL, the effect of co-depletion of *HAVCR1* and *LDLR* on LDL binding was compared to that of single depletion of *HAVCR1*. No significant difference was observed (Fig. 31). Furthermore, the effect of overexpression of GFP-tagged wt*HAVCR1* on LDL uptake in *LDLR*-depleted cells was compared to its effect in control cells. However, the overexpression of wt*HAVCR1* could not rescue the effect of *LDLR* depletion on LDL uptake (Fig. 32). These findings contradict the initial

hypothesis that HAVCR1 might act as an alternative receptor for LDL, in the absence of LDLR. HAVCR1 might rather act as a chaperone for LDLR, thus promoting LDL uptake, as its combinatorial depletion with LDLR further inhibited LDL uptake (Fig. S2.21).

Besides, HAVCR1 has also signaling functions, which are mediated via the tyrosine kinase phosphorylation motif located on its cytoplasmic tail (Fig. 37) (Tietjen et al. 2014). Therefore, an alternative hypothesis could be that HAVCR1 functions as a signaling factor in the LDL internalization pathway.

Moeover, *HAVCR1* demonstrated a synergistic effect with *LDLR* on *LDLR* mRNA (Fig. 22) and protein (Fig. 24), which could suggest that *HAVCR1* might affect the transcriptional and translational regulation of *LDLR*.

Furthermore, HAVCR1 was reported to co-localize and interact in endosomes and lysosomes with NPC1 (Kuroda et al. 2015), the role of which in transporting cholesterol from late-endosomes and lysosomes has already been described (see section 1.3.2.3). Hence, HAVCR1 could take part in the endolysosomal trafficking of LDLR. Specifically, the mechanism of action of HAVCR1 might rather involve later steps of the LDL endocytosis pathway, occurring for example in the endo-lysosomal compartment.

Elucidation of the role of *HAVCR1* in LDL endocytosis and/or trafficking through testing the above hypotheses could moreover give insight into the contribution of this gene to atherosclerosis development, since HAVCR1 has been reported to attenuate atherosclerosis (Hosseini et al. 2015).

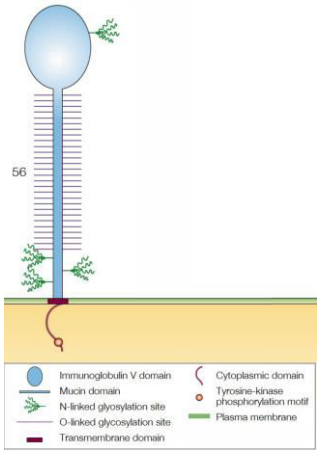


Figure 37. Schematic representation of HAVCR1 protein structure. Glycosylation sites were predicted using NetOglyc and NetNglyc. (Adapted by permission from Macmillan Publishers Ltd: Nature Rev. Immun. Kuchroo et al., copyright 2003)

4.7 HAVCR1 RARE VARIANTS

HAVCR1 was considered as a very interesting interactor in my co-RNAis screen, since its co-depletion with *LDLR* potentiated the effect of *LDLR* depletion on LDL uptake, and it also interacted with a number of other genes. Therefore, its role in LDL endocytosis was further examined, by testing the effect of variants of the protein. In particular, 19 variants were extracted from the Exome Variant Server (http://evs.gs.washington.edu/EVS/), which contains data from patients with extreme lipid traits and/ or atherosclerosis, as well as healthy individuals. Of the 19 variants, 15 are rare (MAF<5%) and four are common (MAF>5%). Surprisingly, overexpression of 18 out of 19 tested variants of *HAVCR1* in HeLa Kyoto cells had a significant inhibitory effect on LDL uptake (Fig. 33, 34). Of those variants, most are localized in the coding sequence of the extracellular domain of the protein; therefore they could affect PtdSer binding. Two variants (K317E, A343E) are localized in the cytoplasmic part, and could affect signal transduction. One variant (Y297C) is localized on the transmembrane part of the protein.

Interestingly, four variants that were predicted as benign by Polyphen-2 (V156I, L179P, T207A and M158T) had a significant inhibitory effect as well, while on the other hand, a variant that was predicted as probably damaging (V138I) had no significant effect.

The finding that almost all tested variants had an effect on LDL uptake could be due to factors not directly related to the uptake of LDL. Noteworthy, five of the variants that significantly decreased LDL uptake (V69G, D99H, N114S, A343E, S100G) had a toxic effect on the cells overall, decreasing the total cell number more than 50%, as compared to cells treated with the wild-type protein. Hence, these variants might interfere with more basic cell functions, such as cell growth, division or survival. However, the remaining ten variants with a significant effect on LDL uptake were not toxic for the cells. Therefore, the inhibition of LDL uptake caused by these variants is probably not attributed to decreased cell viability but rather directly related to endocytosis. These findings strongly suggest a role for HAVCR1 in LDL uptake, which is disturbed by these mutations.

Nonetheless, it should be determined if the effect of HAVCR1 variants is specific to the endocytosis of LDL or is a rather general effect on clathrin-mediated endocytosis.

For this, their effect should be further tested on the internalization of other ligands, such as transferrin or EGFR (epidermal growth factor receptor).

4.8 SNP CO-OCCURENCE

The genes that were tested in the co-RNAi screen are located in close proximity to lead SNPs of loci identified in GWA studies as associated with lipid traits and/ or CAD/MI. The identification of interactions among these genes implies the existence of synergistic combinations among the SNPs, nearby which the genes under study are mapped. Moreover, as many studies have shown that the co-occurrence of SNPs can predict the risk of getting diseases, a question that arises is whether the co-occurrence of SNPs in genes identified to interact will be associated with increased CVD risk.

In order to investigate SNP-SNP interactions and co-occurence, the total 144 lead SNPs for each of the 30 genes that were tested in the co-RNAi screen were tested for pairwise interactions in a cohort of 4893 individuals with subclinical atherosclerosis (Bioimage study) (Muntendam et al. 2010), in collaboration with Heiko Runz (Merck Research Laboratories, US). For the (70) SNPs that were present on the exome chip used for genotyping in the Bioimage study, as well as for the proxies of those not present, single-variant LDL association test was performed by Gulum Kosova, based on the allele-frequency of the LDL-raising allele.

None of the SNP pairs tested showed a statistically significant interaction; one reason for this might be the relatively low size of the study sample. Specifically, the detection of epistatic effects requires measurements taken from very large population sizes, in order to obtain sufficient statistical power (Carlborg and Haley 2004; Phillips 2008). Therefore, in order to bring non-significant detected SNP interactions to statistical significance, larger sample sizes probably need to be tested.

Nonetheless, 36 SNP pairs showed a significant additive effect (see Tbl. S13). Namely, for these 36 pairs it was shown that plasma LDL increased proportionally to the number of LDL-increasing alleles. Interestingly, 3 SNP pairs (*LPL* + *CELSR2*, *APOB* + *HMGCR*, *LDLR*+*NCAN*) that showed an additive effect, corresponded to genes that were identified as interactors for LDL-uptake in the co-RNAi screen (see Fig. 35).

These findings corroborate a synergistic effect for these genes in cholesterol regulation.

Next, SNP co-occurrence was investigated for those interactors of the co-RNAi screen, which were mapped close to SNPs that were present on the Exome Chip, or that had proxies that were present on it (*PCSK9*+*TMEM57*, *LDLR*+*MLXIPL*, *LDLR*+*HAVCR1*, *MLXIPL*+*HAVCR1*), as well as for the triplet *LDLR*+*MLXIPL*+*HAVCR1*. In individuals showing co-occurrence of the corresponding SNPs, the LDL levels increased proportionally to the number of risk alleles. In the case of *MLXIPL*+*HAVCR1*, this was also true for TG levels. Interestingly, for *PCSK9*+*TMEM57* the LDL levels were significantly higher in the case of heterozygosity for both risk alleles (2 risk alleles), as compared to homozygosity for at least one of two risk alleles (3 or 4 risk alleles) (see Fig. 36).

Overall, the SNP interaction/ co-occurence analysis demonstrated that lipid levels, and especially those of LDL strongly correlate with the total number of risk alleles present. Given the fact that elevated LDL levels increase the risk for CVD, the amount of lipid-increasing alleles also correlates with the disease risk.

5 IMPLICATIONS FOR FURTHER RESEARCH

This study investigated genetic interactions underlying cholesterol homeostasis in human, and specifically LDL-cholesterol endocytosis. This was done in an effort to uncover gene networks that control complex polygenic diseases that arise from a deregulation in lipid metabolism, most importantly cardiovascular disease. With this study, genetic interactions were identified between previously unrelated genes, including also genes without a previously described role in cholesterol metabolism. For many of the identified interactions, hypotheses were constructed for the mechanistic interpretation of the synergistic effect of the genes on LDL endocytosis, based on corroborating effects on *LDLR* expression, as well as on the transcription of *SREBF*, the transcription factor of *LDLR*.

However, the exact mechanisms of genetic interactions still need to be elucidated. Further experiments are needed to define if any physical interactions exist between the products of interacting genes, or between transcription factors and promoter sites. For instance, a potential protein-protein interaction between *LDLR* and *HAVCR1* needs to be examined, in order to test the hypothesis of an assisting role for HAVCR1 in LDLR-mediated endocytosis. Moreover, potential interaction between *MLXIPL* transcription factor and *LDLR* or *HAVCR1* should be examined, to test the hypothesis of a role for *MLXIPL* in the expression of the two genes. Furthermore, the effect of the detected gene interactions on different steps in the pathway of LDL endocytosis could be explored, such as the transport of LDLR back to the plasma membrane and the glycosylation of LDLR, which is required for its stable expression at the cell surface. In addition, the effect of gene interactions on the efficiency of the cell transport machinery, from the membrane to the endocytic compartments and from those back to the plasma membrane could be tested. Specifically, effects on COPII-mediated vesicle formation and its components -such as Sec23, Sec24 and Sar1- as well as on recycling endosomes and their components -such Rab11- could be tested.

In parallel to the *in vitro* experiments, this study attempted to uncover SNP-SNP interactions in patients, between SNPs within or in close proximity to genes that were identified to interact. For this, a cohort of nearly 5000 individuals with

subclinical atherosclerosis was genotyped and single-variant association analysis was performed for lipid traits. Statistically significant additive effects were detected between SNPs, but none of the SNP-SNP interactions reached statistical significance. To bring these interactions to statistical significance, two possible approaches could be considered; one is to repeat the analysis in a much larger sample that would increase the chances for epistasis detection. The second approach includes the correlation of the clinical findings with *in vitro* experiments, whereby two supposedly interacting SNPs could be simultaneously introduced, and their synergistic effect on cholesterol homeostasis could be monitored. In the future, *in vivo* experiments in animal models could be also considered, to model the effect of SNP-SNP interactions on cholesterol regulation and atherosclerosis development.

All the approaches and experiments outlined above will increase the data to be assessed for the elucidation of genetic interaction mechanisms. The results of the present study lay the basis for more advanced investigations. A potential approach to a better understanding of the system could be integration of the generated data into a testable and predictive model. In this way, a better understanding of the complex nature of lipid traits and cardiovascular disease could be achieved.

6 MATERIALS

6.1 EUKARYOTIC CELL LINES

Cell line	Tissue	Supplier
HeLa Kyoto	Human cervical carcinoma	S. Narumiya, Kyoto University, Japan (gift to J. Ellenberg, EMBL HD)
HuH7	Human hepatocellular carcinoma	Japanese Collection of Research Bioresources JCRB (#JCRB0403) (Nakabayashi et al., 1982)
HEK293	Human embryonic kidney	ATCC: CRL-1573 Kai Simons, Dresden
Sk-Hep1	Human ascites/ Liver adenocarcinoma	ATCC: HTB-52 Martin Beck Lab, EMBL
HLE	Human hepatoma	Martin Beck Lab, EMBL
HLF	Human lung fibroblasts	Martin Beck Lab, EMBL

6.2 CELL CULTURE MEDIA

Medium	Composition	Supplier
Fetal bovine serum (FBS)		PAA #A15-101
L-glutamin		Sigma Aldrich #G7513
Penicillin 10,000U-Streptomycin 10 mg/mL (Pen-Strep)		Sigma #P0781
DMEM (Dulbecco's Modified Eagle Medium)	With: Low Glucose (1g/L), L-Glutamine, Sodium Pyruvate, Phenol Red	GIBCO #31885-023
Supplemented medium	DMEM 1g/L glucose 10% (v/v) FCS 1% (v/v) L-glutamin	Self produced
Supplemented medium, with antibiotics	DMEM 1g/L glucose 10% (v/v) FCS 1% (v/v) L-glutamin 1% (v/v) P/S	Self produced
Starving medium	DMEM 1g/L glucose 1% (v/v) L-glutamin 0.2% (v/v) BSA	Self produced
OptiMEM |Reduced Serum Medium		GIBCO #31985070
0.05% Trypsin-EDTA		GIBCO #25300054
Freezing medium	DMEM 1g/L glucose 1% (v/v) L-glutamin 20% (v/v) FCS 10% (v/v) DMSO	Self produced
Imaging Medium	30 mM Hepes 115 mM NaCl 1.2 mM CaCl2 1.2 mM MgCl2 1.2 mM K2HPO4 2 g/l D-glucose Medium was adjusted to pH 7.4	EMBL media kitchen

6.3 BUFFERS AND SOLUTIONS

Buffer	Composition	Supplier
1M HEPES (4-(2-hydroxyethyl-)piperazin-1-ethansultonic acid)	23.8% (w/v) HEPES (Biomol #05288.100) in ddH_2O adjusted to pH 7.25, autoclaved	Media kitchen
$MgCl_2$	0.5M $MgCl_2$ in ddH_2O	Self produced
MOPS (1x) running buffer	NuPAGE® MOPS SDS running buffer (20x) Diluted to 1x in ddH_2O	Self produced
PFA 3% fixation buffer	3% (v/v) PFA in PBS	Pepperkok lab
PBS (Phosphate buffer saline)	137 mM NaCl 2.7 mM KCl 1.4 mM KH2PO4 4.8 mM NA2HPO4/Na2CO3 Adjusted to pH 7.4, autoclaved	Media kitchen
Quenching buffer	30 mM glycine	Pepperkok lab
Methanol/Acetone permeabilization buffer	50% Methanol 50% Ethanol stored at -20° C	Self produced
Saponin permeabilization buffer	0.1% (v/v) saponin in PBS	Self produced
Triton X-100 permeabilization buffer	0.1% (v/v) Triton X-100 in PBS	Self produced
Western blot blocking buffer	0.1% (v/v) Tween-20 5% (w/v) milk powder in PBS	Self produced
Tris-HCl 0.5M	Trizma-base 6.06g Adjusted to pH 6.8 with HCl ddH_2O to 100 ml	Self produced
Western blot sample buffer: SDS-loading buffer (2x)	80 mM Tris-HCl 10% Glycerol 4.5% SDS 130 mM DTT 0.005% Bromoph. Blue Diluted in ddH_2O to 10 ml	Self produced
Western blot (mild) stripping buffer for reprobing	15 g glycine 1 g SDS 10 ml Tween20 Add ddH_2O to 1L Adjust pH to 2.2 Bring volume up to 1 L with ultrapure water	Self produced
Western blot (harsh) stripping buffer for reprobing	20 ml SDS 10% 12.5 ml Tris HCl 0.5M pH 6.8 67.5 ml ultra pure water 0.8 ml ß-mercaptoethanol	Self produced
Western blot transfer buffer	25mM Trizma Base 190mM Glycine 20% MeOH for proteins > 80kDa add 0.5% SDS Dilute in ddH_2O to 1 L	Self produced
Western blot wash buffer (PBS-T)	1% (v/v) Tween-20 in PBS	Self produced

6.4 REAGENTS

Reagent	Manufacturer
Ampicillin	Applichem #A0839.0025
Benzonase nuclease, purity >90%	Novagen 25 U /µl #70746
Bovine serum albumin (BSA)	Sigma #A7979
Bromphenol blue Sodium Salt	Applichem #A1120
Dil-LDL	Invitrogen #L3482
Dimethylsulfoxide (DMSO)	Merck Milipore #102952
DRAQ5	Biostatus #DR50200
Drying pearls orange	Aldrich #94098
1.4 Dithiothreitol (DTT)	Sigma Aldrich #D0632-1G
Ethanol	Merck Millipore #100983
Filipin III	Sigma #F4767
Gelatine	Sigma #69391
Gelatine from pig skin, Oregon Green® 488 conjugate	Molecular Probes #G13186
Glycerol 85% pro analysis	Merck #104094
Glycerol 99.0-101.0%	Sigma #15523-1L-R
Glycine: GR for analysis	Merck Millipore #56-40-6
Hoechst, 1mg/ml	Molecular Probes #H33342
(2-Hydroxypropyl)-β-cyclodextrin (HPCD) powder	Sigma #C0926
Hydrochloric acid	Merck Millipore #100317
Lipofectamine 2000	Invitrogen #11668-019
Magnesium chloride hexahydrate	Merck #1058331000
β-mercaptoethanol	Sigma #M1511
Methanol	Merck Millipore #106009
Milk powder	Reform, Granovita GmbH
NuPAGE® MOPS SDS Running Buffer (20x)	Novex #NP0001
Mowiol® 4-88	Calbiochem #475904
Nonidet® P40 (NP-40)	Applichem #A1694
Oligofectamine	Invitrogen #12252011
PageRuler Plus Prestained Protein Ladder	Thermoscientific #26619
Paraformaldehyde (PFA) 16% Solution EM Grade	Electron Microscopy Science #15710
Paraformaldehyde (PFA) powder EM grade	Polysciences #00380-250
Poly-L-lysine hydrobromide	Sigma #P6282
Ponceau S	Applichem #A1405
RNAse H 120U (2U/ml)	Invitrogen #18021-071
Saponin	Sigma Aldrich #47036
Sodium Dodecyl Sulfate (SDS) 20% (w/v)	BioRad #161-0418
Sodium hydroxide	Merck Millipore #109137
Sucrose	Affymetrix, USB # 21938
SYBR Green	Applied Biosystems #4309155
Triton-X-100	Sigma #T9284
Trizma-base	Sigma #T1503
Tween-20	Sigma #P7949
Water HPLC grade	Sigma Aldrich #V270733
Western Blotting Substrate, Pierce® ECL Plus	Thermo Scientific #32132

6.5 PLASMIDS AND OLIGONUCLEOTIDES

6.5.1 PLASMIDS

Plasmid	ORF	Tag	ORF [bp]	Total [bp]	Supplier
EX-EGFP-M03	eGFP	C-eGFP		7033	GeneCopoieia
EX-A0821-M03	LDLR	C-eGFP	2583	8931	GeneCopoieia
EX-Z7459-M03	SORT1	C-eGFP	2496	8844	GeneCopoieia
EX-Z7374-M03	DAB2	C-eGFP	2313	8673	GeneCopoieia
EX-H9327-M03	SEZ6L	C-eGFP			GeneCopoieia
EX-Z9616-M03	CXCL12	C-eGFP	423	6783	GeneCopoieia
EX-Z0263-M03	HAVCR1	C-eGFP	1095	7443	GeneCopoieia
EX-Z0263-M03	HAVCR1-W47R	C-eGFP	1095	7443	Heiko Runz lab.
EX-Z0263-M03	HAVCR1-V69I	C-eGFP	1095	7443	Heiko Runz lab.
EX-Z0263-M03	HAVCR1-V69G	C-eGFP	1095	7443	Heiko Runz lab.
EX-Z0263-M03	HAVCR1-D99H	C-eGFP	1095	7443	Heiko Runz lab.
EX-Z0263-M03	HAVCR1-S100G	C-eGFP	1095	7443	Heiko Runz lab.
EX-Z0263-M03	HAVCR1-R110C	C-eGFP	1095	7443	Heiko Runz lab.
EX-Z0263-M03	HAVCR1-N114S	C-eGFP	1095	7443	Heiko Runz lab.
EX-Z0263-M03	HAVCR1-V138I	C-eGFP	1095	7443	Heiko Runz lab.
EX-Z0263-M03	HAVCR1-T149A	C-eGFP	1095	7443	Heiko Runz lab.
EX-Z0263-M03	HAVCR1-V156I	C-eGFP	1095	7443	Heiko Runz lab.
EX-Z0263-M03	HAVCR1-T174M	C-eGFP	1095	7443	Heiko Runz lab.
EX-Z0263-M03	HAVCR1-L179P	C-eGFP	1095	7443	Heiko Runz lab.
EX-Z0263-M03	HAVCR1-T207A	C-eGFP	1095	7443	Heiko Runz lab.
EX-Z0263-M03	HAVCR1-E223stop	C-eGFP	1095	7443	Heiko Runz lab.
EX-Z0263-M03	HAVCR1-D264N	C-eGFP	1095	7443	Heiko Runz lab.
EX-Z0263-M03	HAVCR1-Y297C	C-eGFP	1095	7443	Heiko Runz lab.
EX-Z0263-M03	HAVCR1-K317E	C-eGFP	1095	7443	Heiko Runz lab.
EX-Z0263-M03	HAVCR1-A343E	C-eGFP	1095	7443	Heiko Runz lab.
EX-Z0263-M03	HAVCR1-M158T	C-eGFP	1095	7443	Heiko Runz lab.

6.5.2 PRIMER OLIGONUCLEOTIDES

Gene	Fwd primer 5´->3´	Rev primer 5´->3´
ACTB	CCACGAAACTACCTTCAACTCC	CTCGTCATACTCCTGCTTGCT
BCAM	GTGCGCTTGTCTGTACCC	ATATAATGGTCGTGGGTTCC
BCL7B	CCCGAGAACCTAATGGCTTT	CGGAACTCTGGTTGCTGTTT
CELSR2	ATGACACGCTCATCTGGAGT	AGGACCTTTCTTCTCAAAGCCC
GAPDH	CATGAGAAGTATGACAACAGCCT	AGTCCTTCCACGATACCAAAGT
HMGCR	CCATCCCTGGGAAGTCATAG	AGGATGGCTATGCATCGTG
HRPT1	TGAGGATTTGGAAAGGGTGTT	CAGAGGGCTACAATGTGATGG
HSPC3	ATGGAAGAGAGCAAGGCAAA	AATGCAGCAAGGTGAAGACA
LDLR	CGATGAAGTTGGCTGCGTTA	GTTGCAGACTTTGTCCAGGG
LDLRAP1	GTCGCCACGGGGAATTAT	TTGTCGTGCATCTTGTCTG
LPL	GTGGCTACCTGTCATTTCAA	GCACCCAACTCTCATACATTC
MLXIPL	CAGCAACAAGACCGAGAACC	GCACTGAGTGTGCTCACGA
PAFAH1B1	GTGTCTGCCTCAAGGGATAA	ACGTACCCATTCTCTGTGTC
PCSK9	GTGAAGATGAGTGGCGACC	GTAATCCGCTCCAGGTTCCA
PVRL2	CCTGATACCTGTGACCCTCT	CCGAGGTACCAGTTGTCATC
RPS18	TGTGGTGTTGAGGAAAGCA	CTTCAGTCGCTCCAGGTCTT
SDHA	TGGTGCTGGTTGTCTCATTA	ACCTTTCGCCTTGACTGTT
SIK3	GGAGCAGGCAGGTACTGCTAT	CTTCACCCTCATCACTGTCCAA
SORT1	ATGGCTATTGGTCCTGAGAA	TTCTTCGCAAAATCTGATGA
SREBF1	CAGCATAGGGTGGGTCAAAT	GAGCCGTGCGATCTGGA

SREBF2	ATGGGCAGCAGAGTTCCTTC	CGACAGTAGCAGGTCACAGG
TMEM57	GAGGGATCCGCTCAGAAATGG	GGCTTCCTGCTCAGCTTTTAG
TOMM40	ATTCAGATGGAGGGTGTCAAG	ATTGTGCTGAGGGCTACTGT

6.5.3 SIRNA OLIGONUCLEOTIDES

Gene Symbol	siRNA ID	Sense siRNA Sequence 5´->3´	Antisense siRNA Sequence 5´->3´
APOB	s1476	GCAAGUACCUGAGAACGGAtt	UCCGUUCUCAGGUACUUGCtg
APOE	s1496	CUAGUUUAAUAAAGAUUCAtt	UGAAUCUUUAUUAAACUAGgg
BAZ1B	s17209	CCUUCGUAGUGAUCUCAUUtt	AAUGAGAUCACUACGAAGGaa
BCAM	s8338	CGACCAUUAUAUGCUGGAAtt	UUCCAGCAUAUAAUGGUCGtg
BCL7B	s228480	AGACAUUGGAGGGAAGGGAtt	UCCCUUCCCUCCAAUGUCUct
CBLC	s24223	GGGAGGCCGUGAGUAUCUAtt	UAGAUACUCACGGCCUCCCag
CELSR2	s4526	GACCCAGACUGCAACAAGAtt	UCUUGUUGCAGUCUGGGUCaa
CXCL12	s445518	CCAUGGAGGCACUAACAAAtt	UUUGUUAGUGCCUCCAUGGca
FAM174A	s445532	CCGCGCCUAUGGUCCCUCUtt	AGAGGGACCAUAGGCGCGGug
HAVCR1	s230290	CGACUGUUCUGACGACAAUtt	AUUGUCGUCAGAACAGUCGtt
HMGCR	s143	GGUUCGCAGUGAUAAAGGAtt	UCCUUUAUCACUGCGAACCct
LDLR	s224006	CAGCGAAGAUGCGAAGAUAtt	UAUCUUCGCAUCUUCGCUGgg
LDLRAP1	s25120	CCAACCAGCUCAUUGAGAAtt	UUCUCAAUGAGCUGGUUGGtg
LPL	s702	GCAACAAUCUGGGCUAUGAtt	UCAUAGCCCAGAUUGUUGCag
MLXIPL	s27387	GCAAGCUGGUGUCUCCCAAtt	UUGGGAGACACCAGCUUGCca
MYBPHL	s50998	UGAACCUACUAAUCCCAUUtt	AAUGGGAUUAGUAGGUUCAct
MYLIP	s26522	GACUUUAGCCCAAUUAAUAtt	UAUUAAUUGGGCUAAAGUCat
NCAN	s3648	GGCCUUCUGUAAACAGGAAtt	UUCCUGUUUACAGAAGGCCat
PAFAH1B1	s445574	GGAGGGACAUACCACUAUAtt	UAUAGUGGUAUGUCCCUCCUU
PAFAH1B2	s10001	GGAACAAAUAACCACGAAAtt	UUUCGUGGUUAUUUGUUCCta
PCSK9	s48694	GGAACAAAUAACCACGAAAtt	atCCUUGUUUAUUGGUGCUUU
PVRL2	s445582	GCAUUUCACCAUUCAAACAtt	UGUUUGAAUGGUGAAAUGCcc
SEZ6L	s24052	GGCUAGAGCGCAGCGGGAAtt	UUCCCGCUGCGCUCUAGCCuc
SIK3	s445595	GCCUGAAAAUGGAUACCAAtt	UUGGUAUCCAUUUUCAGGCag
SORT1	s224557	GAAUGGUCGAGACUAUGUUtt	AACAUAGUCUCGACCAUUCtg
TM6SF2	s28703	GCUUCUUCGUGUGCAAUCUtt	AGAUUGCACACGAAGAAGCag
TMEM57	s30473	GCUUCGGAAAUAUAAGGAAtt	UUCCUUAUAUUUCCGAAGCtc
TOMM40	s20449	CACGCAACAUACUACCACAtt	UGUGGUAGUAUGUUGCGUGca
WDR12	s31441	GACUGGAUCAGUUCAAUUAtt	UAAUUGAACUGAUCCAGUCat
ZNF259	s445636	GCUCUUUCUGUAGGUUAUUtt	AAUAACCUACAGAAAGAGCag
INCENP	s7424	AGUCCUUUAUUAAGCGCAAtt	UUGCGCUUAAUAAAGGACUtc
NPC1	s237198	CCAAUUGUGAUAGCAAUAUtt	AUAUUGCUAUCACAAUUGGtc
neg9	s444246	UACGACCGGUCUAUCGUAGtt	CUACGAUAGACCGGUCGUAtt
neg1	s813	UAACGACGCGACGACGUAAtt	UUACGUCGUCGCGUCGUUAtt

6.6 ANTIBODIES

6.6.1 WESTERN BLOT ANTIBODIES

6.6.1.1 PRIMARY WESTERN BLOT ANTIBODIES

Protein	Host	Supplier	Cat. No.	Clone	Dilution
a-Tubulin	mouse	Neomarkers	#MS581	monoclonal	1:10000
LDLR	rabbit	BioVision	3839-100	polyclonal	1:1000
LDLRAP1/ARH	rabbit	ProteinTech	13213-1-AP	polyclonal	1:1000

6.6.1.2 SECONDARY WESTERN BLOT ANTIBODIES

Protein/Label	Host	Supplier	Cat. No.	Clone	Dilution
HRP-anti-rabbit	goat	Sigma	A0545	polyclonal	1:16000
HRP-anti-mouse	rabbit	Sigma	A9044	polyclonal	1:8000
HRP-anti-goat	mouse	Sigma	A9452	monoclonal	1:10000

6.6.1.3 PRIMARY IMMUNOFLUORESCENCE ANTIBODIES

Protein/Label	Host	Supplier	Cat. No.	Clone	IF Fixation	Dilution
BCAM	mouse	R n D systems	MAB148-SP	monoclonal		1:500
BCAM	rabbit	igma	HPA005654	polyclonal		1:50
CXCL12	mouse	R n D systems	MAB350-SP	monoclonal		1:250
HAVCR1/TIM1	goat	R n D systems	MAB1750	monoclonal	PFA/MeOH	1:250
LDLR	rabbit	Fitzerald	20R-LR002	polyclonal	PFA	1:100
LDLRAP1/ARH	rabbit	ProteinTech	13213-1-AP	polyclonal		1:1000
MLXIPL/CHREBP	goat	SantaCruz	(P-13): sc-21189	polyclonal		1:200
PAFAH1B1/LIS1	mouse	Sigma Aldrich	L7391	monoclonal		1:5000
SORT1/NTR3	rabbit	Sigma	HPA006889	polyclonal		1:100
TOMM40	rabbit	Sigma	HPA036232	polyclonal		1:50

6.6.1.4 SECONDARY IMMUNOFLUORESCENCE ANTIBODIES

Protein/Label	Host	Supplier	Cat. No.	Clone	Dilution
AlexaFluor 488-mouse	goat	Molecular Probes	#A11001	polyclonal	1:200
AlexaFluor 488-rabbit	goat	Molecular Probes	#A11008	polyclonal	1:400
AlexaFluor 488-goat	donkey	Molecular Probes	#A11055	polyclonal	1:400

6.7 KITS

Kit	Supplier	Cat.No.
Western blot		
Invitrap Spin Universal RNA Mini Kit	Stratec Molecular	#1060100300
Superscript III, First-Strand Synthesis Supermix for RT-qPCR	Invitrogen	#11752-050

6.8 EQUIPMENT

6.8.1 MICROSCOPES

Microscope	Manufacturer
Automated widefield microscope Scan^R	Olympus
Axiovert 200 widefield microscope	Zeiss

6.8.2 OTHER LABORATORY EQUIPMENT

Device/Model	Manufacturer
VersArray ChipWriter Pro System	BioRad
Liquidator 96 manual pipetting system 360 Rainin	Mettler-Toledo
MiVac Quattro concentrator, GeneVac	SPScientific

Cell culture incubator	Binder
Centrifuge 5417R	eppendorf
Centrifuge 5804R	eppendorf
Kodak RP X-OMAT Processor M6B	Kodak
Large bench centrifuge Multifuge 3s	Heraeus
Magnetic stirring hotplate MR3001 K	Heidolph
Mastercycler gradient	eppendorf
Microcentrifuge 5417R	eppendorf
Mini-PROTEAN® II Cell Gel System	Bio-Rad #1599
Nanodrop 8000 Spectrophotometer	PeqLab
pH Meter #MP225	Mettler-Toledo
PowerPac Power Supply	BioRad
QuantStudio™ 6 Flex Real-Time PCR System	Applied Biosystems
Scale SBA 51	Scaltec
Scale TE124S-OCE	Sartorius
Scanner Perfection V750 Pro	Epson
Sonificator	Hielscher Ultrasound Technology
StepOne™ Real-Time PCR System #4376357	Applied Biosystems
Thermomixer C	eppendorf
Tube shaker	eppendorf
Vortex Genie2	Scientific Industries
Water bath	GFLR

6.8.3 EQUIPMENT FOR CELL CULTURE AND MICROSCOPY

Equipment	**Supplier/ Cat. No.**
Cell counting chamber	Superior Marienfeld #0610610
Cell culture dish 10 cm	Nunc #150350
6-well cell culture plate	Nunc #140675
12-well cell culture plate	Falcon #353225
24-well cell culture plate	Falcon #353226
96-well glass bottom plate	Zellkontakt #21315241
96-well black bottom plate	Nunc #137101
96-well deep well plate	Greiner #780270
96-well conical bottom microwell plate	Nunc #277143
384-well glass bottom plate	BD Falcon #353962
384-well low volume plate	Nunc #264360
384-well deep well plate	Eppendorf #0030 521.145
Cell freezing container	Nalgene #5100-0001
Cell scrapers	Corning® #3010
Chemiluminescent films	Amersham Hyperfim™ ECL #28906835
Chemiluminescent films	Kodak
Coverslips 11 mm diameter	Menzel-Gläser #CB00110RA1
Coverslips 15 mm diameter	Menzel-Gläser #CB00150RA1
Cryotubes	Nunc #363401
Immobilon® PVDF membrane	Merck Millipore #ISEQ00010
LabTek 1 well	Nunc #155361
MicroAmp® Fast Optical 96-well Reaction Plate with Barcode	Applied Biosystems #4346906
MicroAmp® Optical 384-well Reaction Plate with Barcode	Applied Biosystems #4326270
MicroAmp® Optical Adhesive Film	MicroAmp #4311971
NuPAGE® 4-12% Bis-Tris Gel 1.0mm x10 well	NOVEX #NP0321BOX
NuPAGE® 4-12% Bis-Tris Gel 1.0mm x12 well	NOVEX #NP0322BOX

NuPAGE® 4-12% Bis-Tris Gel 1.0mm x15 well	NOVEX #NP0323BOX
Object Slides frosted end	Menzel-Gläser #ISO 8037/I
Parafilm®	BEMIS® #PM996
Sterile filter 0.45 µm	Millipore Stericup Durapore™ #SCHVU05RE
Tips for Liquidator	Steinbrenner #SL-LT-L200
Whatmann Paper	Whatman® #3030917
Water reservoir	Thermo Fisher #370906
X-Ray cassette	KISKER #IEC 60406
Combitips advanced	eppendorf #0030089774

6.9 SOFTWARE

Software	Developer
Adobe Acrobat X Pro	Adobe systems Inc., San Jose, USA
Adobe Acrobat Reader DC	Adobe systems Inc., San Jose, USA
Adobe Illustrator CS6	Adobe systems Inc., San Jose, USA
Adobe Photoshop CS6	Adobe systems Inc., San Jose, USA
ApE 2-A plasmid Editor	M. Wayne Davis, biologylabs, Utah
AxioVision	Carl Zeiss
CellProfiler	Broad Institute, Cambridge, USA
EndNote X6	Thomson Reuters, New York, USA
FiJi	Schindelin J. et al.
HTM (High Throughput Microscopy) Explorer	Tischer C., ALMF, EMBL
ImageJ 1.46r	Rasband W., NIH, Bethesda, USA
Mendeley Desktop 1.16.1	Elsevier
Microsoft Office 2011	Microsoft Corporation, Redmond, USA
QuantStudio™ Real-Time PCR Software	Applied Biosystems, USA
R 3.2.2 GUI 1.66	The R Foundation for Statistical Computing, Vienna, Austria
RStudio	RStudio Team (2015)
Scan^R	Olympus, Hamburg, Germany
StepOne Software v2.3	Applied Biosystems, USA

6.10 WEBTOOLS AND RESOURCES

Tool	Usage	Website
BLAST/BLAT	Align sequence to genome/ oligonucleotide design	http://www.ensembl.org/Multi/Tools/Blast
Bluegecko	Information on siRNAs and genome-wide screens	https://bluegecko.embl.de (EMBL intranet only)
BLOCK-iT RNAi Designer	siRNA design	https://rnaidesigner.thermofisher.com/rnaiexpress
Confetti; a multi-enzyme map of the HeLa proteome	Information on protein expression in HeLa cell line	https://proteomics.swmed.edu/confetti/
Exome Variant Server	Rare variant database	http://evs.gs.washington.edu/EVS/
GeneOntology	Information on gene function/ classification	http://www.geneontology.org/
Life Technologies Fluorescence Spectra Viewer	Fluorescence Spectra Viewer	https://www.thermofisher.com/de/de/home/life-science/cell-analysis/labeling-chemistry/fluorescence-spectraviewer.html
PhenoScanner	SNPs database-Cardiovascular Epidemiology	http://www.phenoscanner.medschl.cam.ac.uk/phenoscanner

Primer-BLAST NCBI	PCR primers design	https://www.ncbi.nlm.nih.gov/tools/primer-blast/
Primer 3	PCR primers design	http://bioinfo.ut.ee/primer3-0.4.0/primer3/
PrimerBank	PCR primer database	https://pga.mgh.harvard.edu/primerbank/
qPrimerDepot	PCR primer database	https://primerdepot.nci.nih.gov/
Reverse Complement	Oligonuclotide sequence transformation	http://www.bioinformatics.org/sms/rev_comp.html
Sigma Aldrich OligoEvaluator	Sequence Analysis Tool	http://www.oligoevaluator.com/OligoCalcServlet
SNPcheck	PCR primers design	https://secure.ngrl.org.uk/SNPCheck/snpcheck.htm
STRING	Functional proteins associations network	http://string-db.org/
UCSC In-silico PCR	PCR primers design	http://rohsdb.cmb.usc.edu/GBshape/cgi-bin/hgPcr

7 METHODS

7.1 CELL BIOLOGY

7.1.1 CELL CULTURE

HeLa Kyoto cells were cultured in 10 cm petri dishes (Nunc) with supplemented, low glucose (1g/L) DMEM growth medium (GIBCO) at 37°C, 5% CO_2 and saturated humidity. The cells were passaged every two or three days, once they reached approximately 80% confluence, until passage 24, after which the cells were discarded and fresh cells were thawed (see 7.1.1.2). For passaging, the cells were first washed once with trypsin-EDTA (GIBCO) and then incubated shortly at 37°C with 1 mL trypsin to detach from the plastic dish. Once detached, the cells were suspended in 9 mL supplemented DMEM medium and part of the cell suspension was transferred to a new dish and complemented with medium. For transfection experiments, cells were diluted [1:2] one day prior to plating (see 7.1.2). For the co-RNAi screen, penicillin 100µg/ml- streptomycin/0.2% (w/v) (Sigma) was added to the medium throughout all the experiments.

7.1.1.1 PLATING CELLS

Depending on the type and the duration of the experiment, a certain amount of cells was plated on LabTeks or multi-well plates (see Tbl. 10), after counting with a Hemocytometer (Superior Marienfeld). For experiments performed using multi-well plates in which the cells would be finally imaged, cells were plated on top of glass coverslips placed inside the plate wells.

Table 10. Amounts of cells used in different cell culture plates and times of transfection

Dish format	DMEM volume per well	Cell No	Transfection duration (siRNA/ cDNA)	Total experimental time from plating to end point
1-well LabTek	1.5 mL	$6x10^4$	48 h	48 h
6-well plate	2 mL	$3x10^4$	48 h	72 h
12-well plate	1 mL	$1.5x10^4$	48 h	72 h
12-well plate	1 mL	$4x10^4$	24 h	48 h

7.1.1.2 Freezing and Thawing Cells

The cells were frozen and stored in liquid nitrogen to keep stocks for future use, after being tested for mycoplasma. For this purpose, cells were diluted [1:2] (see 7.1.1) in a 10 cm dish one day prior to freezing, to reach about 80% confluency. The cells were trypsinised (see 7.1.1) and suspended in supplemented medium, and the suspension was centrifuged for 5 minutes at 1,000 rpm. After sedimentation of the cells, the medium was removed, and the cells were suspended in 2 mL freezing medium. Aliquots of 1 mL cell suspension were transferred into labeled cryotubes (Nunc) and stored at -80°C in a freezing container which allowed for slow freezing of the cells. Within the next days, the cells were transferred to the liquid nitrogen tank.

To thaw cells, cryotubes were removed from the liquid nitrogen and quickly thawed by holding shortly (1-2 minutes) in the 37°C waterbath. After thawing, the cells were transferred into a Falcon tube, where they were suspended in supplemented DMEM, before centrifuging for 5 minutes at 1,000 rpm. The supernatant was aspirated, and the sedimented cells were suspended in supplemented DMEM and transferred into a 10 cm dish.

7.1.2 Transfection

In order to increase or inhibit the expression of a gene of interest, nucleic acids were introduced into cells by transfecting them with cDNA or siRNA, respectively. Two different transfection protocols were used, described as forward or liquid-phase transfection and reverse or solid-phase transfection.

7.1.2.1 Forward (Liquid-Phase) Transfection

The conventional transfection protocol employed is referred to as “forward transfection”, whereby cells are seeded one day prior to transfection. For both cDNA and siRNA transfection experiments, two different mixtures were first prepared. The first mixture contained the nucleic acid -cDNA or siRNA- diluted in OptiMEM (GIBCO), whereas the second mixture contained the transfection reagent diluted in OptiMEM. The two mixtures were mixed after incubation at room temperature (RT) for 5 minutes, and incubated for another 20 minutes to allow for the formation of the transfection complexes. The transfection mixture was then added drop-wise to the cells. For cDNA transfection, the transfection mixture was added to cells grown

in supplemented DMEM medium. For siRNA transfection, the cell medium was aspirated, the cells were washed once with serum-free growth medium, and then serum-free DMEM medium was added to the cells before the addition of the transfection mixture. The medium of siRNA-transfected cells was replaced after 4 h incubation with supplemented DMEM medium. The amounts of transfection reagent and nucleic acid differed depending on the method and format used. Typically used amounts are shown in Tables 11 and 12, however these amounts were subject to changes if transfection efficiency was not satisfactory.

Table 11. Typical amounts of siRNA and transfection reagents used

Format	Experiment	siRNA mix (1)	Transfection reagent mix (2)	Final Mix (1+2)
6-well plate	48h siRNA knockdown	30 pmol (1 µL of 30 µM) siRNA in 200 µL OptiMEM	2 µL Lipofectamine in 200 µL OptiMEM	403 µL into 800 µL serum-free DMEM
12-well plate	48h siRNA knockdown	15 pmol (0.5 µL of 30 µM) siRNA in 100 µL OptiMEM	1 µL Lipofectamine in 100µL OptiMEM	201.5 µL into 400 µL serum-free DMEM

Table 12. Typical amounts of cDNA and transfection reagents used

Format	Experiment	cDNA mix (1)	Transfection reagent mix (2)	Final Mix (1+2)
12-well plate	48h siRNA knockdown	1.25µg cDNA in 125µL OptiMEM	3µL Lipofectamine in 125µL OptiMEM	253µL into 1mL supplemented DMEM

7.1.2.2 REVERSE (SOLID-PHASE) TRANSFECTION

For high-throughput knockdown experiments, reverse transfection was used, whereby freshly passaged cells -passaged 24 h before- were added to transfection complexes in pre-spotted microarrays; “LabTeks” (see Tbl. 10).

7.1.2.2.1 LABTEK SPOTTING

For reverse transfection experiments, glass-bottomed chambered cell culture slides (1-well LabTeks) coated with siRNAs for solid phase reverse-transfection of cells were produced as described (Erfle et al. 2007, 2008). In detail, the transfection mixtures were prepared as follows.

First, 0.2 g gelatin (Sigma) was dissolved in 100 ml ddH_2O at 56°C for 20 minutes in a prewarmed waterbath. The gelatin solution was sterilized by filtration with a 0.45 µm pore filter after cooling down. Besides the normal gelatin solution, a 0.2% Oregon Green gelatin solution (Molecular Probes) was prepared by dissolving 5 mg

Oregon Green gelatin in 2.5 mL ddH_2O as previously described. At the same time, 1.37 g sucrose (Affymetrix) was dissolved in 10 mL OptiMEM.

Second, four numbered 96-well conical bottom plates containing 15 pmol Silencer Select siRNA (ThermoFisher) (5 µL of a 3µM solution) per well were prepared and centrifuged at 500 rpm for 1 minute. Additionally, the transfection mix was prepared by mixing 1.662 mL of the sucrose /OptiMEM solution with 0.969 mL ddH_2O and 0.969 mL Lipofectamine (Invitrogen). The transfection mix and the gelatin solution were distributed into two 96-well conical bottom plates, by transferring 32 µL and 50 µL per well, respectively, and the two plates were centrifuged. The gelatin solution was then distributed into four numbered 96-well plates, by transferring 10 µL per well and the plates were centrifuged. The normal gelatin was replaced with Oregon Green gelatin in those wells of the four gelatin plates that would be used for the transfection control siRNA spotting.

Third, 7 µL from the transfection mix plate were transferred to each of the four siRNA plates with the Liquidator and mixed by pipetting. The plates were then incubated at RT for 20 minutes to allow for the formation of the transfection complexes. Afterwards, 7 µL of the four gelatin plates were transferred to the corresponding siRNA/transfection-mix-containing plates with the Liquidator and mixed by pipetting and the plates were centrifuged shortly.

Fourth, 16 µL from the four siRNA transfection cocktail-containing plates were transferred to a low-volume 384-well plate with the Liquidator, and the plate was centrifuged shortly. For better spotting results, LabTeks were washed before spotting with ddH_2O, then ethanol and then again ddH_2O. Finally, the 384-well plate was used to spot LabTeks using a ChipWriter Compact Robot (Bio-Rad) with solid pins (Point Technologies) resulting in a spot volume of ~4 nL (containing ~5 ng siRNA) and a spot diameter of ~400 µm and a spot-to-spot distance of 900 µm, mirroring one 384-well plate per chamber. siRNA arrays were printed in 12 replicates and dried for 10 minutes in a SpeedVac pre-warmed to 37°C. The LabTeks were then stored in sealed plastic boxes containing drying pearls (Aldrich).

7.1.2.2.1.1 Determining the optimal siRNA concentration for the double knockdowns

In order to determine the optimal siRNA concentration to be used for the double knockdowns, two conditions were examined in a pilot experiment. For this purpose, siRNAs targeting two genes with a known role in cholesterol uptake were used (LDLR, DAB2), as well as a negative control siRNA (NEG9) and a transfection control siRNA (INCENP).

In detail, LabTeks were spotted with the following combinations of siRNAs; Neg9+Neg9, LDLR+LDLR, DAB2+DAB2, LDLR+NEG9, DAB2+NEG9, LDLR+DAB2, INCENP+INCENP using either 15pmol of each siRNA (30pmol siRNA in total) or 7.5pmol of each siRNA (15pmol siRNA in total) per LabTek spot. The LDL uptake was performed on the spotted LabTeks as described (see 7.1.6.1) and the knockdown effect of different siRNA concentrations on LDL uptake was evaluated. The experimental settings are shown schematically in Table 13.

Table 13. siRNA amount used in the pilot experiment to determine optimal concentrations for double knockdowns

siRNA combination	Amount of each siRNA	Total siRNA amount
Neg9+Neg9	15pmol (2.5µL of 6µM)	30pmol (5 µL of 3µM)
LDLR+LDLR	15pmol (2.5µL of 6µM)	30pmol (5 µL of 3µM)
LDLR+LDLR	7.5pmol (2.5µL of 3µM)	15pmol (5 µL of 1.5µM)
DAB2+DAB2	15pmol (2.5µL of 6µM)	30pmol (5 µL of 3µM)
DAB2+DAB2	7.5pmol (2.5µL of 3µM)	15pmol (5 µL of 1.5µM)
LDLR+NEG9	15pmol (2.5µL of 6µM)	30pmol (5 µL of 3µM)
LDLR+NEG9	7.5pmol (2.5µL of 3µM)	15pmol (5 µL of 1.5µM)
DAB2+NEG9	15pmol (2.5µL of 6µM)	30pmol (5 µL of 3µM)
DAB2+NEG9	7.5pmol (2.5µL of 3µM)	15pmol (5 µL of 1.5µM)
LDLR+DAB2	15pmol (2.5µL of 6µM)	30pmol (5 µL of 3µM)
LDLR+DAB2	7.5pmol (2.5µL of 3µM)	15pmol (5 µL of 1.5µM)
INCENP+INCENP	15pmol (2.5µL of 6µM)	30pmol (5 µL of 3µM)
INCENP+INCENP	7.5pmol (2.5µL of 3µM)	15pmol (5 µL of 1.5µM)

7.1.2.2.1.2 Spotting of LabTeks for the Combinatorial RNAi Screen

For the combinatorial RNAi screen, LabTeks coated with siRNAs were produced as described previously (see 7.1.2.2.1). For the spotting of two siRNAs per LabTek spot, the four 96-well conical bottom plates contained 15 pmol of each of the two Silencer Select siRNAs (2 x 2.5 µL of a 6µM solution=30 pmol total siRNA) per well.

Previously validated siRNAs against three genes were selected as controls: INCENP,

inducing multinucleated arrested cells that served to monitor siRNA transfection efficiency (Neumann et al. 2006); LDLR, reducing DiI-LDL uptake to ~25% and perinuclear Filipin signal to ~60%; NPC1, increasing perinuclear Filipin signal to ~150% (Bartz et al. 2009).

The total 30x30 (900 including the reciprocal interactions) pairwise combinations to be tested in the combinatorial RNAi screen including all the controls were printed on 5 different LabTeks. On each LabTek, the single knockdowns (siRNA(X)+Neg9) were spotted 2 times-with the exception of LDLR+Neg9, that was spotted 10 times, while the double knockdowns (siRNA(X)+siRNA(Y)) were spotted 1 time. Of the 384 spots printed on the LabTeks, 56 spots contained control siRNAs; 8 spots contained INCENP-siRNA (s7424) to control for transfection efficiency, 8 spots contained the non-silencing control siRNA; Neg1 (s229174), 8 spots contained the non-silencing control siRNA; Neg9 (s444246), 8 spots contained siRNA targeting LDLR (s224006) as a positive control for LDL-uptake, 8 spots contained siRNA targeting NPC1 (s237198) as a positive control for free cholesterol (FC), 8 spots contained both Neg9 and LDLR siRNA, as a positive control for the double knockdowns in LDL-uptake assay and 8 spots contained both Neg9 and NPC1 siRNA, as a positive control for the double knockdowns in FC.

7.1.3 IMMUNOSTAINING

For immunofluorescence experiments, cells were seeded at a concentration of $1x10^4$ cells/mL in 12-well dishes containing a glass coverslip in each well. After washing once with PBS, cells were fixed with either paraformaldehyde (PFA) or methanol, depending on the protein and the antibody. For PFA fixation, cells were incubated for 20 minutes at RT in 3% PFA and then for 5 minutes with quenching buffer. For methanol fixation, cells were incubated with methanol at -20°C for 4 minutes. Following fixation, cells were washed three times for 3 minutes with PBS at RT. For intracellular protein labeling, cells needed to be permeabilized with either Triton-X-100 buffer for 5 minutes at RT, or saponin permeabilization buffer for 10 minutes at RT. Cells were then washed three times with PBS. For the staining, the antibody was diluted in PBS or in saponin buffer in case of cells permeabilized with saponin. The cells were incubated with the primary antibody, then washed three times with PBS

and then incubated with the secondary antibody. Subsequently, cells were washed with PBS and their nuclei were stained with Hoechst (diluted 1:5,000) for 5 minutes at RT. After washing once more with PBS, coverslips were mounted on glass slides using Mowiol. The glass slides were dried overnight at RT or at 37° C for at least 30 minutes prior imaging.

7.1.4 Microscopy-Detection of Fluorescent Molecules in the Cell

7.1.4.1 Wide-Field Microscopy

Immunostained cells were imaged with a Zeiss Axiovert 200, inverted, wide-field microscope. The light source was an HBO lamp and the fluorescence filters used were the following; Dapi/Hoechst (Excitation(Ex): 365/12, Emission (Em): LP397), FITC (Ex: 450-490, Em: 515-565), Cy3 (Ex: 545/25, Em: 605/70). The images were acquired with a Zeiss EC Plan-NEO FLUOAR 40x/NA 1.3 Oil or a Zeiss Plan-APOCHROMAT 63x/NA 1.4 Oil objective, using the AxioVision software.

7.1.4.2 High-Throughput Microscopy

For high-throughput microscopy the automated wide-field Olympus Scan^R microscope was used. The light source was a MT20 Xenon lamp and typically an Olympus UPlanSApo 20x/ NA 0.75 objective was used. The fluorescence filters used were the following; Dapi/Hoechst (Ex: 350/50, Em: 447/60), Cy3 (Ex: 545/25, Em: 605/70), Cy5 (Ex: 640/30, Em: 690/50) and GFP (Ex: 470/40, Em: 525/50). A software autofocus was based on the nuclear (Dapi/Hoechst) staining. It consisted of a coarse autofocus (± 13µm, 18 steps) and a fine autofocus (± 2 µm, 15 steps). The starting position of each coverslip/ LabTek, as well as the spot distance (for LabTeks) was set, and the imaging was then performed in an automated manner. In order to find the first spot on a LabTek, the fluorescence signal from the Oregon Green gelatin was used (see 7.1.2.2.1), which was only contained in spots with the transfection control siRNA -including the upper left (A1) spot of the LabTek. Usually 100 images were obtained per coverslip, or 1 image per spot of a LabTek. The imaging settings were stored in an experiment descriptor .xml file, which could be re-used in the following imaging set.

7.1.5 Endocytosis Assays

7.1.5.1 LDL Uptake Assay

To analyse the ability of cells to endocytose low-density lipoprotein (LDL), the LDL-uptake assay was employed (Bartz et al. 2009; Gilbert et al. 2009; Pitas et al. 1981). Cells were seeded in appropriate format and usually transfected with siRNAs for 48 h before performing the assay. Cells were depleted from sterols in two sequential steps, before the assay (see section 7.1.6). They were subsequently washed with ice-cold imaging medium (IM)/0.2% BSA for 2 minutes and then labeled with 1.25µg/mL DiI-LDL (Invitrogen) for 30 minutes at 4 °C. The endocytosis of DiI-LDL was then stimulated for 20 minutes at 37 °C and then cells were washed with IM/0.2% BSA for 2 minutes at 4 °C. To remove membrane-bound DiI-LDL, cells were washed for exactly 1 minute with IM at pH 3.5, and then again with IM/0.2% BSA for 2 minutes at 4 °C. Fixation of cells in 3% PFA for 20 minutes followed and after washing with PBS, cells were counterstained for cytoplasm and nucleus with DRAQ5 and Hoechst, for 5 minutes at RT, respectively. Finally, cells were washed with PBS and either imaged directly in the case of LabTeks, or mounted with Mowiol onto glass slides, in the case of coverslips.

7.1.5.2 LDL Binding Assay

To test for the ability of siRNA-treated versus non-treated cells to bind LDL, a similar cellular assay to the LDL-uptake assay was performed. The difference in the LDL-binding assay was that the cells were only let to bind LDL for 30 minutes at 4° C, with no LDL-internalization step following. Immediately after the binding step, cells were washed 3 times with IM-BSA and then fixed with PFA.

7.1.6 Sterol-Depletion Protocols

Cells were depleted from sterols prior to the LDL-uptake (or LDL-binding) assay, in two steps. For this purpose, 28 h after the transfection (20 h before the assay) the cells were washed once with serum-free DMEM medium and then serum-free DMEM medium containing 0.2% BSA was added to the cells. Additionally, 45 minutes before the assay, 1% (w/v) HPCD was given to the cells to extract cholesterol for the medium.

7.2 MOLECULAR BIOLOGY

7.2.1 RNA EXTRACTION AND REVERSE TRANSCRIPTION

In order to measure gene expression on mRNA level, total RNA needs to be extracted from cultured cells. For this purpose, 30000 HeLa Kyoto cells/ well were seeded in 6-well plates three days before the extraction. 24 hours after seeding, cells were transfected with siRNAs and 48 hours after the transfection RNA was extracted using the Invitrap Spin Universal RNA Mini Kit (Stratec), according to manufacturer's instructions. The concentration of the extracted RNA was quantified by measuring its absorbance at 260 nm using the Nanodrop 8000 Spectrophotometer. Samples were either directly used for reverse transcription or stored at -20° C until usage. Similarly, the RNA purity was assessed by measuring the absorbance ratio A260/A280 nm. Total extracted RNA was then reverse-transcribed into cDNA using the SuperScript™ III First-Strand Synthesis SuperMix for RT-qPCR (Invitrogen). According to manufacturer's instructions, 500 ng RNA were used as template for a 10 µl total volume reaction. cDNA samples were wither directly used for quantitative real-time polymerase chain reaction (RT-qPCR) or stored at -20° C until usage. Before usage, cDNA samples were diluted 1:10 with RNAse-free water.

7.2.2 QUANTITATIVE REAL-TIME POLYMERASE CHAIN REACTION (QRT-PCR)

RT-qPCR was performed using SYBR Green fluorescent dye (Applied Biosystems). The -relative or absolute- quantity of the amplified DNA is thereby calculated by measuring the SYBR Green fluorescence at the end of each amplification cycle. In each well of a MicroAmp® optical 96-well reaction plate (Applied Biosystems), 2 µl of the previously synthesized cDNA were added, as well as 18 µl of a mix consisting of 11 µl SYBR Green, 1.1 µl forward and 1.1 µl reverse primer (10 µM) and 6.6 µl ddH_2O, to a total volume of 20 µl. For each gene target, different mixes were prepared. All reactions were performed in triplicates within the same plate. The plate was sealed with a MicroAmp® Optical Adhesive Film and centrifuged at 1000 rpm for 2 minutes.

RT-qPCR experiments were performed at the GeneCore Facility at EMBL using the StepOne™ Real-Time PCR System (Applied Biosystems). The experimental setup used was as follows: experiment type "Quantitation-Comparative C_T ($\Delta\Delta C_T$)", reagent

type “SYBR® Green Reagents”, ramp speed “Standard (~2 hours to complete a run)” and the following program was run: (1) hot start polymerase activation at 95 °C for 10 minutes, (2) 40 cycles of dissociation at 95 °C for 15 seconds and elongation at 60 °C for 1 minute, (3) primer melt curve at 95 °C for 15 seconds, 60 °C for 1 minute, a gradient of +0.3 °C to 95 °C and finally 95 °C for 15 seconds. Upon completion of the run, RT-qPCR data were analyzed with the StepOne Software v2.3 (Applied Biosystems) using the comparative C_T method.

7.2.3 HIGH-THROUGHPUT QRT-PCR

For measuring the expression of multiple genes in multiple samples simultaneously on the same plate, MicroAmp® Optical 384-well Reaction Plates (Applied Biosystems) were used and the RT-qPCR was run on a QuantStudio™ 6 Flex Real-Time PCR System (Applied Biosystems). For the generation of the 384-well plates, four MicroAmp® Optical 96-well reaction plates were produced and combined with a Liquidator into one MicroAmp® Optical 384-well Reaction Plate. In each of the four 96-well plates, 2 µl of the synthesized cDNA were pipetted, as well as 8 µl of a mix consisting of SYBR Green, forward (10 µM) and reverse primer (10 µM) and ddH_2O. For one well mix, 2.2 µl ddH_2O, 0.55 µl FW primer (10µM), 0.55 µl RV primer (10µM) and 5.5 µl Sybr Green 1 were combined, and different mixes were prepared for each gene target. Thus, the total reaction volume per well was 10 µl. The four 96-well plates were combined onto the 384-well plate using a Liquidator. The 384-well plate was then centrifuged, and the RT-qPCR reaction was run using the same settings as in the StepOne™ Real-Time PCR System. The RT-qPCR was analyzed using the QuantStudio® Real-Time PCR Software v1.3.

7.2.4 PRIMER VALIDATION FOR QRT-PCR

Primer pair oligonucleotides were custom designed for the qRT-PCR experiments (see section 7.4.5.2). Before use, primer pairs were evaluated for their binding to the target gene, using cDNA synthesized from total RNA extracted from untreated cells. Four serial dilutions of cDNA were prepared by diluting with a serial factor {1:10}; 1:1, 1:10, 1:100, 1:1000, which were used as standards to construct a standard curve, in order to estimate the PCR amplification efficiency of each primer. Thus, the

concentration of the “standards” was indicated as 1, 0.1, 0.01 and 0.001. Aside the “standards”, two “negatives” were used; a no template control (NTC) which omits the cDNA and serves as a control for extraneous nucleic acid contamination, and a no reverse transcriptase control (NRT), which serves as a control for DNA contamination in the RNA preparation. All standards and negatives were used in duplicate wells, to assess variation in C_q (quantitation cycle). The slope of the standard curve and the binding efficiency were evaluated, and primers were rejected when they didn’t give efficient results; slope range between -3.58 and -3.10 and efficiency between 90% and 110% were considered acceptable.

7.2.5 Selection of Reference Gene for qRT-PCR experiments

In order to compare mRNA transcription through qRT-PCR in different samples or tissues, the appropriate reference gene (aka “housekeeping gene”) needs to be selected, for normalization purposes. Five candidate reference genes were tested; GAPDH (Glyceraldehyde 3-Phosphate Dehydrogenase), HRPT1 (Hypoxanthine Phosphoribosyltransferase 1), RPS18 (Ribosomal Protein S18), SDHA (Succinate Dehydrogenase Complex Flavoprotein Subunit A) and HSPC3 (Small Heat-Shock Protein 3) for their expression stability under different experimental settings, which was evaluated using qRT-PCR.

For this purpose, total RNA was extracted from HeLa Kyoto cells, untransfected or transfected and cDNA was synthesized from the extracted RNA. Primer sequences for HRPT1, RPS18, SDHA and HSPC3 were taken from Krainova et al. (Krainova et al. 2013). All primer sequences are shown in section 6.5.2.

7.2.6 Nucleid acids Handling

Small interfering RNAs (siRNAs) were delivered lyophilized in tubes or plates in a concentration of 5nM, and were reconstituted by resuspending in nuclease-free H_2O to a final concentration of 30µM. siRNAs were then used at this concentration in liquid-phase transfections, or were further diluted in nuclease-free H_2O to 3µM for the reverse transfections. Plates and tubes with siRNAs were stored at -20 °C. Lyophilized oligonucleotide primers were reconstituted by resuspending in ddH_2O to

100 µM stock solutions, which were stored at -20 °C. For qRT-PCR primer 100 µM stock solutions were diluted to 10 µM working solutions with ddH_2O.

7.3 BIOCHEMISTRY

7.3.1 SODIUM DODECYL SULFATE POLYACRYLAMIDE GEL ELECTROPHORESIS (SDS-PAGE)

To assess the amount of a protein of interest in cells, cells were grown in 6-well plates and after washing once with PBS they were lysed in 200µL sample buffer on ice. After transferring to a 1.5 mL eppendorf tube, cell lysates were boiled at 95 °C for 5 minutes. After brief spinning and cooling down of the samples, 1% (v/v) benzonase as well as 1% (v/v) $MgCl_2$ were added to the each tube and let to degrade DNA at RT for 10-20 minutes. Afterwards, samples were boiled for 5 minutes at 95 °C to ensure complete protein denaturation, and either stored at -20 °C or were used directly after cooling down. For SDS-PAGE, 18 µL of the samples were loaded on precast 12-well 4%-12% NuPAGE Bis-Tris gels (Invitrogen), and run using MOPS buffer at 100-140 V. To determine the molecular weight of the proteins, 4 µL protein ladder was loaded onto the first well of the gel.

7.3.2 WESTERN BLOT

After separating with SDS-PAGE, proteins were transferred to a PVDF membrane, where they would later be detected using specific antibodies. Initially, a PVDF membrane piece was activated in methanol for 10 minutes. Whatmann papers and Western blot sponges were also soaked in cold transfer buffer. The cassette was assembled using in the following order from the white to the black back of the cassette; sponge, two Whatmann papers, SDS-PAGE gel, PVDF membrane, two Whatmann papers, sponge. The cassette was then placed in the blotting chamber, which was filled with cold transfer buffer and an ice pack and a magnetic stirrer. The transfer of the proteins from the gel to the PVDF membrane was then performed in the cold-room for 1 h at 400mA.

After the transfer, the PVDF membrane was immersed into blocking buffer and placed on a shaker for at least 1 h at RT. The membrane was then incubated with the primary antibody in blocking buffer shaking overnight in the cold room. Next day, the membrane was washed with PBS-T for 1-2 h at RT, before incubation with the

secondary antibody diluted in blocking buffer, which was performed at RT for 1 h on a shaker. After washing three times for 10 minutes with PBS-T, a chemiluminescence reaction was performed to detect the protein signal, using the Pierce® ECL Plus Western Blotting Substrate. For this, Substrate A and Substrate B were mixed at a 40:1 ratio and incubated on a membrane for 5 minutes at RT. After the chemiluminescence reaction, the blot was placed between two plastic foils inside of a developing X-ray cassette and the signal was detected in the darkroom by exposing the membrane to a photographic film (Kodak). The films were developed in a Kodak developing machine and scanned using an Epson scanner, in order to be quantified with ImageJ.

7.3.3 STRIPPING OF WESTERN BLOT MEMBRANES

In order to reprobe a membrane with a different antibody, the blotted PVDF membrane was stripped to remove the bound antibodies, with a mild stripping buffer made according to Abcam protocol. The membrane was washed 3 times for 10 minutes in fresh stripping buffer, then 2 times for 10 minutes in PBS, and finally 2 times for 5 minutes in PBS-T. After placing in blocking buffer for at least 1 h, membrane was ready for incubation with the primary antibody.

7.4 COMPUTATIONAL BIOLOGY

7.4.1 IMAGE ANALYSIS

7.4.1.1 IMAGEJ

Images acquired with the Axiovert microscope were imported in ImageJ and inspected visually for fluorescence intensity.

Images acquired with the Scan^R widefield microscope from Olympus were initially visually inspected using ImageJ, in order to exclude from the analysis any out-of-focus images or otherwise non-analyzable.

7.4.1.2 CELLPROFILER

7.4.1.2.1 LDL-UPTAKE ASSAY ANALYSIS

To quantify images from the LDL-uptake assay, images were analyzed with CellProfiler v2.1.1, using a pipeline developed with the help of Volker Hilsenstein at

ALMF, EMBL Heidelberg. For this purpose, the images were first converted from 12-bit images to 16-bit, using the TransferTool (V. Hilsenstein). In the CellProfiler pipeline (see Tbl. 14), image sets that consisted of three channels (Dapi for Hoechst, Cy3 for DiI-LDL and Cy5 for DRAQ5) were first loaded (*LoadImages*) and pixel intensity values were rescaled between 0 and 0.0624856 (*Rescale Intensity*). The intensity rescaling was done to convert 12-bit images saved in 16-bit format to the correct range; the value of 0.0624856 is equivalent to 2^{12}divided by 2^{16}. Then, image quality was measured using the Cy3 channel that reflected the DiI-LDL signal (*MeasureImageQuality*). Here, the image blur was measured and blur metrics were calculated with special algorithms (eg. PowerLogLogSlope). Afterwards, the image intensity of the DiI-LDL channel was measured (*MeasureImageIntensity*), by measuring pixel intensity. In the next step, the nuclei of the cells were identified, using the Dapi/Hoechst signal (*IdentifyPrimaryObjects*), after setting a threshold range for the fluorescence intensity, as well the diameter range of the nuclei. Then, the whole cell area as well as the cell boundaries were identified using the DRAQ5 fluorescence signal, by digitally expanding the nuclei identified with the previous module, after setting a threshold and a threshold correction factor (*IdentifySecondaryObjects*). The cytoplasm was expanded by 15 pixels in the next module, in order to perform the background subtraction (*ExpandOrShrinkObjects*). In the next module, a mask is created from the DiI-LDL channel, by inverting the mask found in the previous module from the cytoplasm, in order to measure the image background intensity (*MaskImage*), which is measured in the next module (*MeasureImageIntensity*). Afterwards, the background fluorescence intensity is subtracted from the DiI-LDL intensity (*ImageMath*). Then, cells touching the image borders are discarded (*IdentifySecondaryObjects*). In the next two steps, a threshold is applied to find saturated pixels in the image, and a binary imaged is generated (*ApplyThreshold*) and the intensity of saturated pixels in the binary image is measured, to define the number of saturated pixels per cell (*MeasureObjectIntensity*). Next, cells with more than a specified number of saturated pixels (typically 10) in the DiI-LDL channel are discarded (*FilterObjects*). The DiI-LDL signal is afterwards enhanced over the cellular background signal (*EnhanceOrSuppressFeatures*), and the fluorescence signal parameters of identified

DiI-LDL dots are measured (*MeasuredObjectIntensity*). Then, DiI-LDL dots outside of cells are excluded (*Crop*) and dots within cells are finally identified (*IdentifyPrimaryObjects*), after setting a threshold. Next, a mask is created that consists of the dots that were previously identified (*MaskImage*), and the size and shape features of the identified dots are measured (Measure*ObjectSizeShape*), as well as the fluorescence intensity of the dots within whole cells from the background corrected DiI-LDL image and the enhanced signal image (*MeasureObjectIntensity*). In the following two modules, a binary image is generated out of the identified DiI-LDL dots (*ConvertObjectsToImage*) and dots are assigned to the cells which they belong to (*RelateObjects*). In the next three modules, the mean total intensity of DiI-LDL for one cell (*CalculateMath*), the total area of DiI-LDL dots in one cell (*CalculateMath*) and the ratio of the DiI-LDL dots intensity to the total intensity (*CalculateMath*) are calculated. In the following 16 modules, images are generated which visualize the outlines of the segmented objects on a black image and the fluorescence intensity data are shown on the objects and the images are saved as .tif files (*OverlayOutlines*, *DisplayDataOnImage, SaveImages*). Finally, the results are exported and saved as .csv files (*ExportToSpreadSheet*).

Table 14. LDL-uptake assay pipeline

1. LoadImages	23. ConvertObjectsToImage
2. Rescale Intensity	24. RelateObjects
3. Rescale Intensity	25. CalculateMath
4. MeasureImageQuality	26. CalculateMath
5. MeasureImageIntensity	27. CalculateMath
6. IdentifyPrimaryObjects	28. OverlayOutlines
7. IdentifySecondaryObjects	29. OverlayOutlines
8. ExpandOrShrinkObjects	30. OverlayOutlines
9. MaskImage	31. OverlayOutlines
10. MeasureImageIntensity	32. DisplayDataOnImage
11. ImageMath	33. DisplayDataOnImage
12. IdentifySecondaryObjects	34. DisplayDataOnImage
13. ApplyThreshold	35. DisplayDataOnImage
14. MeasureObjectIntensity	36. DisplayDataOnImage
15. FilterObjects	37. SaveImages
16. EnhanceOrSuppressFeatures	38. SaveImages
17. MeasureObjectIntensity	39. SaveImages
18. Crop	40. SaveImages
19. IdentifyPrimaryObjects	41. SaveImages
20. MaskImage	42. SaveImages
21. MeasureObjectSizeShape	43. SaveImages
22. MeasureObjectIntensity	44. ExportToSpreadSheet

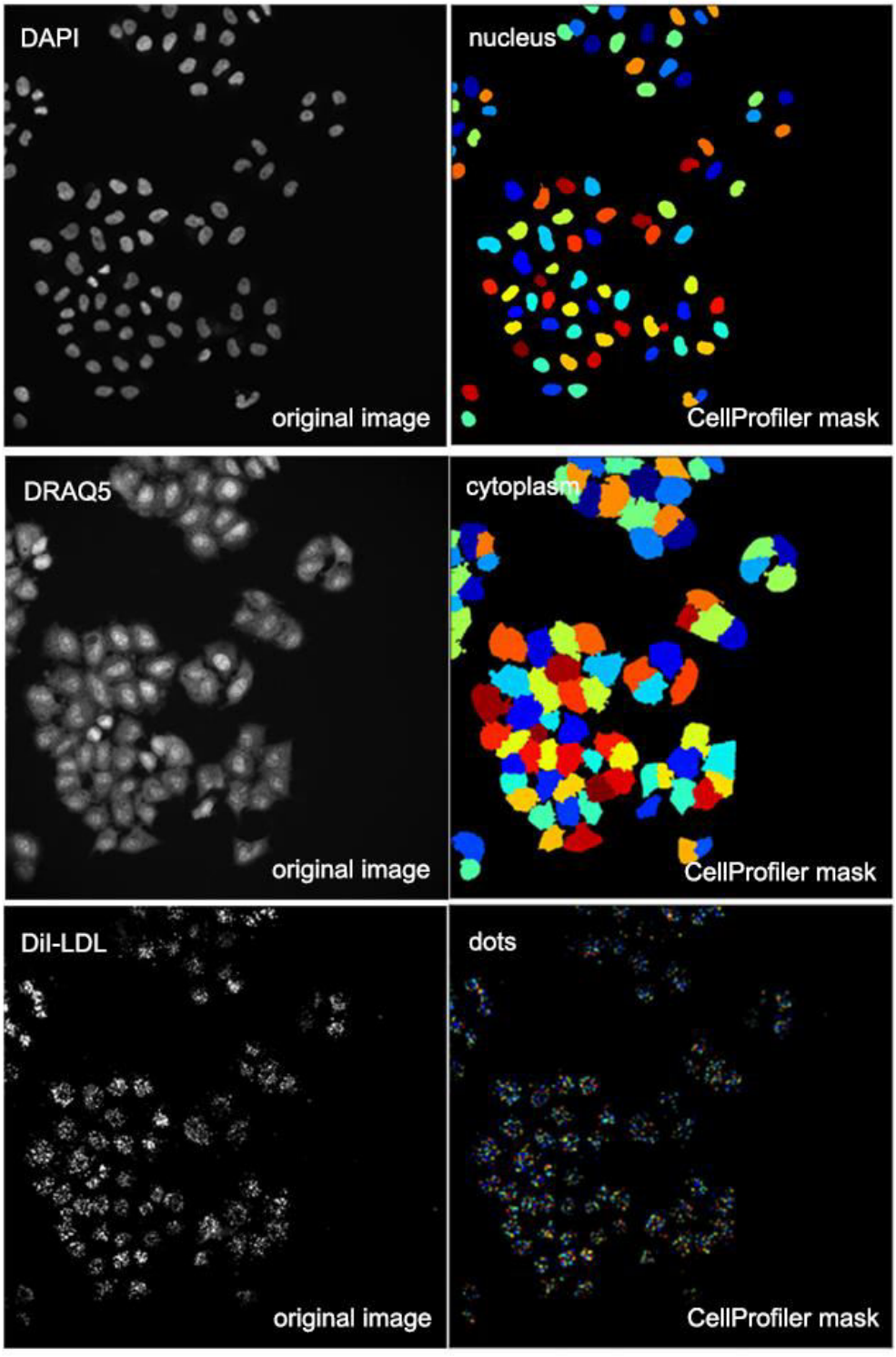


Figure 38. Cellprofiler pipeline for quantitative image analysis of LDL-uptake assay. Shown are representative images from the LDL-uptake assay (left column) acquired with the Scan^R software of the Olympus widefield microscope with the 20x objective and the segmentation of cellular structures (right column) performed through a Cellprofiler pipeline.

7.4.1.2.2 GFP-TAGGED VARIANT OVEREXPRESSION ANALYSIS

In order to quantify the effect on LDL uptake of the overexpression of GFP-tagged HAVCR1 variants, two CellProfiler pipelines were sequentially used, after the initial visual inspection and filtering of images with ImageJ. In the first pipeline, the GFP threshold for transfected cells was determined, that would be used in the second pipeline, which calculated the effect of the cDNA transfection on LDL uptake.

In the first pipeline (Tbl. 15), only the images from untransfected cells were loaded (*LoadImages*), and the fluorescence intensity was rescaled in the Cytoplasm, DAPI and GFP channels (*RescaleIntensity*). The image quality on the DAPI channel was measured (*MeasureImageQuality*) and the cell nuclei (*IdentifyPrimaryObjects*) and cytoplasm (*IdentifySecondaryObjects*) were identified as described previously (see 7.4.1.2.1). Then, the cells were digitally expanded by 15 pixels (*ExpandOrShrinkObjects*) and the expanded cytoplasms were identified (*IdentifySecondaryObjects*). A mask was created by applying on the GFP image the inverted signal of the cell cytoplasm (*MaskImage*) and the GFP intensity of the cells was measured (*MeasureImageIntensity*). Afterwards, the GFP signal of the cells was normalized to the backround of the GFP image (*ImageMath*) and the cell GFP intensity was measured (*MeasureObjectIntensity*). Finally, the results were exported on a .csv spreadsheet (*ExportToSpreadSheet*). The pipeline modules are shown below.

Table 15. Pipeline for the identification of the GFP threshold

1. LoadImages	8. ExpandOrShrinkObjects
2. RescaleIntensity	9. IdentifySecondaryObjects
3. RescaleIntensity	10. MaskImage
4. RescaleIntensity	11. MeasureImageIntensity
5. MeasureImageQuality	12. ImageMath
6. IdentifyPrimaryObjects	13. MeasureObjectIntensity
7. IdentifySecondaryObjects	14. ExportToSpreadSheet

After exporting the results, the GFP intensity of the 95th percentile of the untransfected cells was calculated with Excel; 4*PERCENTILE(array,0.97) using the measurement; *Mean_UnfilteredCytoplasm_Intensity_MeanIntensity_GFPbgCorr.* This number was then used in the second pipeline.

In the second pipeline (Tbl. 16), all the other images -excluding the ones from untransfected cells- were loaded as described previously, and the intensities in all

four channels (DAPI, cytoplasm, DiI-LDL, GFP) were rescaled. Then, steps 5-11 were performed as described before, and the dimensions of the enlarged cytoplasms were measured (*MeasureObjectSizeShape*). In the next step, the cells were filtered based on their size (*FilterObjects*), to exclude very small cells that could be apoptotic. Subsequently, a mask was created on the LDL channel and the GFP channel to measure the fluorescence background (*MaskImage*), and the background intensities were measured (*MeasureImageIntensity*). Next, the LDL background and GFP background were subtracted from the corresponding images (*ImageMath*). Afterwards, a threshold was applied to find saturated pixels in the LDL image (*ApplyThreshold*) and the number of saturated pixels was measured (*MeasureObjectIntensity*). In the following step, the cells with more that 10 saturated pixels in LDL channel were filtered out, and the GFP threshold from the previous pipeline was applied to filter transfected cells (*FilterObjects*). The DiI-LDL signal was afterwards enhanced over the cellular background signal (*EnhanceOrSuppressFeatures*), and the fluorescence signal parameters of identified DiI-LDL dots are measured (*MeasuredObjectIntensity*). Then, DiI-LDL dots outside of cells are excluded (*Crop*) and dots within cells are finally identified, after setting a threshold (*IdentifyPrimaryObjects*). Next, a mask is created that consists of the dots that were previously identified (*MaskImage*), and the size and shape features of the identified dots are measured (*ObjectSizeShape*), as well as the fluorescence intensity of the dots within whole cells from the background corrected DiI-LDL image and the enhanced signal image (*MeasureObjectIntensity*). In the following two modules, a binary image is generated out of the identified DiI-LDL dots (*ConvertObjectsToImage*) and dots are assigned to the cells which they belong to (*RelateObjects*). In the next three modules, the mean total intensity of DiI-LDL for one cell (*CalculateMath*), the total area of DiI-LDL dots in one cell (*CalculateMath*) and the ratio of the DiI-LDL dots intensity to the total intensity (*CalculateMath*) are calculated. In the following 16 modules, images are generated which visualize the outlines of the segmented objects on a black image and the fluorescence intensity data are shown on the objects and the images are saved as .tif files (*OverlayOutlines*, *DisplayDataOnImage, SaveImages*). Finally, the results are exported and saved as .csv files (*ExportToSpreadSheet*).

Table 16. Pipeline for the analysis of Variant Overexpression effect on LDL uptake	
1. LoadImages	29. MeasureObjectIntensity
2. RescaleIntensity	30. ConvertObjectsToImage
3. RescaleIntensity	31. RelateObjects
4. RescaleIntensity	32. CalculateMath
5. RescaleIntensity	33. CalculateMath
6. MeasureImageQuality	34. CalculateMath
7. MeasureImageIntensity	35. OverlayOutlines
8. IdentifyPrimaryObjects	36. OverlayOutlines
9. IdentifySecondaryObjects	37. GrayToColor
10. ExpandOrShrinkObjects	38. OverlayOutlines
11. IdentifySecondaryObjects	39. OverlayOutlines
12. MeasureObjectSizeShape	40. OverlayOutlines
13. FilterObjects	41. DisplayDataOnImage
14. MaskImage	42. DisplayDataOnImage
15. MaskImage	43. DisplayDataOnImage
16. MeasureImageIntensity	44. DisplayDataOnImage
17. ImageMath	45. DisplayDataOnImage
18. ImageMath	46. SaveImages
19. ApplyThreshold	47. SaveImages
20. MeasureObjectIntensity	48. SaveImages
21. FilterObjects	49. SaveImages
22. OverlayOutlines	50. SaveImages
23. EnhanceOrSuppressFeatures	51. SaveImages
24. MeasureObjectIntensity	52. SaveImages
25. Crop	53. SaveImages
26. IdentifyPrimaryObjects	54. SaveImages
27. MaskImage	55. ExportToSpreadSheet
28. MeasureObjectSizeShape	

7.4.1.3 HTM (HIGH-THROUGHPUT-MICROSCOPY) TOOL

The image data generated from CellProfiler software were imported into HTM Explorer, an R-based software developed by C. Tischer from EMBL Heidelberg, for visual inspection, quality control and statistical analysis. Thresholds were set for the cell number range (5 to 150 cells per image), as well as for the quality (using the *PowerLogLogSlope* measurement from CellProfiler) and for the background of the DiI-LDL image. Finally, the data were normalized to the negative control and the measurements are saved.

7.4.2 ANALYSIS OF SCREENING RESULTS

Screening results were analysed using R Studio, R version 3.2.2. In the R code created by the author of this manuscript, the HTM output files of all the LabTeks analyzed in the screen were read one after the other, and the intensities were normalized to the negative control (*Neg9*). To achieve this, for each plate, the mean

robust z score per image, as well as the mean median absolute deviation, MAD of the image were calculated. In detail, the robust z score of an image was calculated by subtracting from the fluorescence signal (mean intensity of DiI-LDL dots per cell) from an siRNA treatment ("*treated*"), the median fluorescence signal of all the negative control siRNAs across the same plate ("*median(controls)*") and dividing by the MAD of these controls ("mad(controls)");

$$robustZscore = \frac{treated - median(Ctrls)}{mad(Ctrls)}.$$

Then, a median robust z score per treatment was calculated for each treatment, by taking the median of robust z scores of same treatments across different biological replicates. The R script is shown below with some basic descriptions of the code.

```
#######
# Scipt written by Anthi Trasta
#Combinatorial RNAi screen with LDL-uptake assay
# Calculation of robust Z scores for each treatment, by normalizing to controls

# Preparations
rm(list=ls()) ### clear R-memory
library(xlsx) ### load xlsx package
setwd("/Volumes/t2pepperkok/Anthi/ScreenAnalysis/HTM_ImageTables")
### set data path for MacOSX
flnms<-list.files(path=".")
flNr<-length(flNms)
Tblall<-data.frame

# Import data and make calculations
for (i in 1:flNr){
    Tbl<-read.csv(flNms[i]) ###read table and save into vector named Tbl
    trueTbl<-subset(Tbl,Tbl$HTM_qcImages==TRUE)&(Tbl$Mean_Cytoplasm_Math_DiITotalIntensityDots_byMean!="NA"))
    ### select rows that correspond to images which passed the quality control and where the
    ### measurement to be analyzed is not NA
    trueTbl<-subset(trueTbl,select=c("Metadata_gene_sorted","Mean_Cytoplasm_Math_DiITotalIntensityDotsbyMean"))
    ###select columns to be analyzed
    trueTbl$Metadata_gene_sorted<-as.factor(trueTbl$Metadata_gene_sorted) ###set treatment as factor
    trueTbl$Mean_Cytoplasm_Math_DiITotalIntensityDots_byMean<-
    as.numeric(trueTbl$Mean_Cytoplasm_Math_DiITotalIntensityDots_byMean) ###set measurement as numeric
    medianTbl<-
    aggregate(trueTbl$Mean_Cytoplasm_Math_DiITotalIntensityDots_byMean,by=list(trueTbl$Metadata_gene_sorted),
    FUN="median",na.rm=TRUE) ###calculate median of each treatment and save in a table named medianTbl
    names(medianTbl)<-c("treatment","medianDiIintensity") ###name columns of medianTbl
    medianCtrl<-medianTbl$medianDiIintensity[medianTbl$treatment=="XWNeg9__NA"]
    ###calculate median of controls of the plate
    MADTbl<-
    aggregate(trueTbl$Mean_Cytoplasm_Math_DiITotalIntensityDots_byMean,by=list(trueTbl$Metadata_gene_sorted),
    FUN="mad",na.rm=TRUE) )
    ###calculate median absolute deviation of each treatment and save in a table named MADTbl
    names(MADTbl)<-c("treatment","MADofDiIintensity") ###name columns of MADTbl
    MADctrl<-MADtbl$MADofDiIintensity[MADtbl$treatment=="XWNeg9__NA"] ###calculate MAD of controls of the plate
    trueTbl$robustZscore=(trueTbl$Mean_Cytoplasm_Math_DiITotalIntensityDots_byMean-medianCtrl)/MADctrl
    ###calculate robust z score per treatment
    Tblall<-rbind(Tblall,trueTbl) ###each time this for-loop is run, the table generated is merged with a combined table, which
    ###in the end consists of all the rows of all the tables generated
```

```
}
robustZscoreTbl<-aggregate(Tblall$robustZscore,by=list(Tblall$Metadata_gene_sorted),FUN="median",na.rm=TRUE)
###calculate robust Z score per treatment
names(robustZscoreTbl)<-c("treatment","robustZscore") ###name columns of robustZscoreTbl
madRobZscoreTbl<- aggregate(Tblall$robustZscore,by=list(Tblall$Metadata_gene_sorted), FUN="mad",na.rm=TRUE)
###calculate mad of each robust Z score
names(madRobZscoreTbl)<-c("treatment","madRobZscore")  ###name columns of madRobZscoreTbl
robustZscoreTbl<-merge(robustZscoreTbl,madRobZscoreTbl,by=c("treatment")) ###merge the two tables
names(robustZscoreTbl)<-c("treatment"," robustZscore"," madRobZscore")

#Save Results
wb<-createWorkbook() ###create new workbook
Sheet1<-createSheet(wb,sheetName="Tblall") ###create new sheet in the workbook
addDataFrame(x=Tblall,sheet=Sheet1,row.names=F) ###save Tblall in Sheet1
Sheet2<-createSheet(wb,sheetName="robustZscoreTbl") ###create new sheet in the workbook
addDataFrame(x=robustZscoreTbl,sheet=Sheet2,row.names=F) ###save robustZscoreTbl in Sheet2
saveWorkbook(wb,"/Volumes/t2pepperkok/Anthi/ScreenAnalysis/Tblall.xlsx") ###save workbook as excel file
write.csv(Tblall,"/Volumes/t2pepperkok/Anthi/ScreenAnalysis/Tblall.csv",row.names=T)
### save Tblall as .csv file to be used in next script
write.csv(robustZscoreTbl,"/Volumes/t2pepperkok/Anthi/ScreenAnalysis/robustZscoreTbl.csv",row.names=T)
###save robustZscoreTbl as .csv file to be used in next script
```

In an alternative approach, the robust Z scores were calculated by subtracting from the fluorescence signal (mean intensity of DiI-LDL dots per cell) of an siRNA treatment ("*treated*"), the median fluorescence signal of all treatments (after excluding the transfection controls; Incenp, and the positive controls; Ldlr and Npc1) across the same plate ("*median(All)*") and dividing by the MAD of all treatments ("*mad(All)*");

$$robustZscore = \frac{treated - median(All)}{mad(All)}$$ (Birmingham et al. 2009; Malo et al. 2006).

Then, a median robust Z score per treatment was calculated for each treatment, by taking the median of robust Z scores of same treatments across different biological replicates. The R scipt is shown below with some basic descriptions of the code.

```
#######
# Scipt written by Anthi Trasta
#Combinatorial RNAi screen with LDL-uptake assay
# Calculation of robust Z scores for each treatment, by normalizing to all treatments

# The preparations are same as in previous script

# Import data and make calculations
for(i in 1:flNr){
    Tbl<-read.csv(flNms[i]) ### read file and save to vector Tbl
    trueTbl<-subset(Tbl,(Tbl$HTM_qcImages==TRUE) &(Tbl$Mean_Cytoplasm_Math_DilTotalIntensityDots_byMean!="NA"))
    ### select rows that correspond to images which passed the quality control and where the
    ### measurement to be analyzed is not NA
    trueTbl<-subset(trueTbl,select=c("Metadata_gene_sorted","Mean_Cytoplasm_Math_DilTotalIntensityDots_byMean",
   "Mean_Cytoplasm_Math_Area_Endos_Mean","Mean_Dots_Intensity_IntegratedIntensity_MaskedLDLTopHat"))
    ###select columns to be analyzed
    trueTbl$Metadata_gene_sorted<-as.factor(trueTbl$Metadata_gene_sorted) ###set treatment as factor
    trueTbl$Mean_Cytoplasm_Math_DilTotalIntensityDots_byMean<-
```

```
as.numeric(trueTbl$Mean_Cytoplasm_Math_DilTotalIntensityDots_byMean) ###set measurement as numeric
########## calculate medians #########
medianTbl<-trueTbl
medianTbl<-subset(medianTbl,(medianTbl$Metadata_gene_sorted!="INCENP__NA")&
(medianTbl$Metadata_gene_sorted!="LDLR__NA")&(medianTbl$Metadata_gene_sorted!="LDLR__NEG9")&
(medianTbl$Metadata_gene_sorted!="NPC1__NA")&(medianTbl$Metadata_gene_sorted!="NEG9__NPC1"))
medianTbl$Metadata_gene_sorted<-"any" ###remove treatment names, for calculation of median
medianAll<-aggregate(medianTbl$Mean_Cytoplasm_Math_DilTotalIntensityDots_byMean,
by=list(medianTbl$Metadata_gene_sorted),FUN="mean",na.rm=TRUE) ###calculated medians of treatments
medianAll<-medianAll$x ###calculate median of all treatments
########## calculate mad #########
MADTbl<-trueTbl
MADTbl<-subset(MADTbl,(MADTbl$Metadata_gene_sorted!="INCENP__NA")&
(MADTbl$Metadata_gene_sorted!="LDLR__NA")&(MADTbl$Metadata_gene_sorted!="LDLR__NEG9")&
(MADTbl$Metadata_gene_sorted!="NPC1__NA")&(MADTbl$Metadata_gene_sorted!="NEG9__NPC1"))
MADTbl$Metadata_gene_sorted<-"any" ###remove names of treatments
MADAll<-aggregate(MADTbl$Mean_Cytoplasm_Math_DilTotalIntensityDots_byMean,
by=list(MADTbl$Metadata_gene_sorted),FUN="mad",na.rm=TRUE) ###calculate medians of treatments
MADAll<-MADAll$x ###calculated mad of all treatments
######### calculate Z score #######
trueTbl$minusMedianAll=(trueTbl$Mean_Cytoplasm_Math_DilTotalIntensityDots_byMean-medianAll)
trueTbl$robustZscore=(trueTbl$Mean_Cytoplasm_Math_DilTotalIntensityDots_byMean-medianAll)/MADAll
write.csv(trueTbl, file = paste("/Users/anthi/ PhD/Experiments/2016/ScreenAnalysis /",flNms[i]),row.names=FALSE)
Tblall<-rbind(Tblall,trueTbl) ### merge each file (Tbl) to a new one (Tblall)
}
robustZscoreTbl<-aggregate(Tblall$robustZscore,by=list(Tblall$Metadata_gene_sorted),FUN="median",na.rm=TRUE)
###calculate robust Z score per treatment
names(robustZscoreTbl)<-c("treatment","robustZscore") ###name columns of robustZscoreTbl
madRobZscoreTbl<- aggregate(Tblall$robustZscore,by=list(Tblall$Metadata_gene_sorted), FUN="mad",na.rm=TRUE)
###calculate mad of each robust Z score
names(madRobZscoreTbl)<-c("treatment","madRobZscore") ###name columns of madRobZscoreTbl
robustZscoreTbl<-merge(robustZscoreTbl,madRobZscoreTbl,by=c("treatment")) ###merge the two tables
names(robustZscoreTbl)<-c("treatment"," robustZscore"," madRobZscore")

### Save Results; this part is same as in previous script
```

7.4.3 Interaction Analysis

7.4.3.1 primary approach; control-based normalization

To test for gene-gene interactions among the results of the combinatorial RNAi screen, the effect on LDL uptake of the combined transfection with two siRNAs targeting two different genes A and B was compared to the additive effect of the two single transfections. In single transfections the respective siRNAs were co-transfected together with the negative control siRNA. For this, a robust linear model was fitted using R, where the additive effect of two single knockdowns A and B {($Gene_A$+control)+($Gene_B$+control)} was regressed to the respective double knockdown effect ($Gene_A$+$Gene_B$), assuming the negative control as an intercept, to derive a common effect estimate for each combination (Cordell 2002). Thus, the interaction effect w_{AB} between two genes A and B was calculated from the difference between the main effects m_A and m_B of the two single knockdowns and the effect of the double knockdown y_{AB} (w_{AB}= m_A+m_B-y_{AB}-y_0), with y_0 being the baseline value, which is given from the negative control siRNA effect. A p-value was

then calculated from the t-value of the linear regression model using the following equation; p_{Val}=2-2*pnorm(abs(t_{Val})). To correct for multiple comparison, the p-value was adjusted using the false discovery rate (fdr) method (Benjamini and Hochberg 1995), and the significance threshold for gene interactions was set at $pValue_{fdr}<10^{-2}$. The R script for the interaction calling is shown below, with descriptions of the steps.

```
#######
# Script written by Anthi Trasta with the help of Bernd Klaus from Huber group
#Combinatorial RNAi screen with LDL-uptake assay
#Interaction calling

#Preparations
rm(list=ls()) ### clear R-memory
library(xlsx) ### load xlsx package
library(MASS) ### load MASS package for rlm function
library(dplyr) ###load dplyr package
setwd("/Volumes/t2pepperkok/Anthi/ScreenAnalysis") ### set data path for MacOSX

#Import data
Tblall<-read.csv ### read Tblall.csv file generated from previous R-script
robustZscoreTbl<-read.csv ### read robustZscoreTbl generated from previous R-script

#Split treatments into two gene names
for (i in 1:nrow(robustZscoreTbl){
    geneName<-unlist(strsplit(as.character(robustZscoreTbl$treatment[i],"__"))
    robustZscoreTbl$Gene1[i]=geneName[1] ###create new column with gene1 name
     robustZscoreTbl$Gene2[i]=geneName[2] ###create new column with gene2 name
}

#Prepare for rlm fitting
robustZscoreTbl$treatment<-as.character(robustZscoreTbl$treatment)
robustZscoreTbl$treatment[robustZscoreTbl$Gene1=="XWNeg9"]<-"NEG9__NEG9"
#rename treatment XWNeg9__NA into NEG9__NEG9
robustZscoreTbl$Gene1[robustZscoreTbl$treatment=="NEG9__NEG9"]<-"NEG9"
robustZscoreTbl$Gene2[robustZscoreTbl$treatment=="NEG9__NEG9"]<-"NEG9"
robustZscoreTbl<-subset(robustZscoreTbl,(robustZscoreTbl$Gene2!="NA"))
### exclude all controls; INCENP, NPC1, NEG1, LDLR, etc
geneTbl<-read.csv("/Volumes/t2pepperkok/Anthi/ScreenAnalysis/geneNames.csv") ###read table with names of 30 genes
geneTbl2<-read.csv("/Volumes/t2pepperkok/Anthi/ScreenAnalysis/geneNamesPlusNEG.csv")
###read table containing the names of the 30 genes plus NEG9
for (i in 1:nrow(Tblall)){
    geneName<-unlist(strsplit(as.character(Tblall$Metadata_gene_sorted[i]),"__"))  #Separate gene names in table Tblall
    Tblall$Gene1[i]=geneName[1] ### create new col in Tblall with gene1
    Tblall$Gene2[i]=geneName[2] ### create new col in Tblall with gene2
    Tblall$treatment[i]<-paste(Tblall$Gene1[i],"__",Tblall$Gene2[i])
}

#Fit robust  linear model
rlmTbl<-data.frame()
for(i in 1:nrow(geneTbl)) {
    testGene1<-geneTbl$gene[i]
    for(j in 1:nrow(geneTbl)){
        testGene2<-geneTbl$gene[j]
        if (testGene1!=testGene2){
        ### create new data frame, with only those combinations from "Tblall" that contain the two testGenes;
        ###Gene1__NEG9,Gene2__NEG9, Gene1__Gene2 and XWNeg9_NA
        ### the "testTbl" is overwritten in each run of the loop
        testTbl<subset(Tblall,(Gene1=="XWNeg9")|((Gene1==testGene1)&(Gene2=="NEG9"))|
                                ((Gene1==testGene2)&(Gene2=="NEG9"))|
                                ((Gene1==testGene1)&(Gene2==testGene2))|
                                ((Gene1=="NEG9")&(Gene2==testGene1))|
                                ((Gene1=="NEG9")&(Gene2==testGene2))|
                                ((Gene1==testGene2)&(Gene2==testGene1)))
        for(k in 1:nrow(testTbl)){
```

```
            if (testTbl$Gene1[k]=="NEG9"){
              testTbl$Gene1[k]<-testTbl$Gene2[k]
               testTbl$Gene2[k]<-"NEG9"
            }
        }
        combi<-(((testTbl$Gene1==testGene1)&(testTbl$Gene2==testGene2))|
                  ((testTbl$Gene1==testGene2)&(testTbl$Gene2==testGene1)))
        ### combi: testGene1_testGene2, testGene2_testGene1
        single1<-((testTbl$Gene1==testGene1)&(testTbl$Gene2=="NEG9")|combi)
        ### single1: testGene1_NEG9, testGene1_testGene2, testGene2_testGene1
        single2<-((testTbl$Gene1==testGene2)&(testTbl$Gene2=="NEG9")|combi)
        ### single2: testGene2_NEG9, testGene1_testGene2, testGene2_testGene1
        rlmfit<-rlm(robustZscore~single1+single2+combi,data=testTbl)
        rs <- summary(rlmfit)$sigma
        rlmfit.df<-as.data.frame(coef(summary(rlmfit)))
        rlmfit.df$Gene1<-paste(testGene1)
        rlmfit.df$Gene2<-paste(testGene2)
        rlmTbl<-rbind(rlmTbl,rlmfit.df)
         }
    }
}
names(rlmfit.df)[2]<-"Std.Error"
names(rlmTbl)[3]<-"tvalue" ###name the 2nd and 3rd column of table rlmTbl
wb<-createWorkbook() ###create new workbook to save results
rlmsheet<-createSheet(wb, sheetName="rlm") ###create new sheet in the workbook
addDataFrame(x=rlmTbl,sheet=rlmsheet, row.names=F) ###add results of robust linear model to new sheet
rlmTbl$coefficient=rownames(rlmTbl) ###create new column in rlmTbl, "coefficient", copying the column "rownames"
rlmTblcombi<-subset(rlmTbl,grepl("combiTRUE",rlmTbl$coefficient))
### create new Table: "rlmTblcombi", with only those rows from rlm Tblall that contain "combiTRUE" which corresponds to the
###value of the combined treatment
rlmTblcombi<-within(rlmTblcombi, rm("coefficient")) ### remove column "coefficient" from rlmTblcombi
for (i in 1:nrow(rlmTblcombi)){
 for (j in 2:nrow(rlmTblcombi)){
   if (rlmTblcombi$Gene1[i]==rlmTblcombi$Gene2[j]){
     if (rlmTblcombi$Gene2[i]==rlmTblcombi$Gene1[j]){
       rlmTblcombi$Gene1[j]<-rlmTblcombi$Gene1[i]
       rlmTblcombi$Gene2[j]<-rlmTblcombi$Gene2[i]
     }
   }
 } ### rename Gene2_Gene1 to Gene1_Gene2
 rlmTblcombi$treatment[i]<-paste(rlmTblcombi$Gene1[i],"__",rlmTblcombi$Gene2[i])
 ### create new column "treatment" with Gene1__Gene2
}
rlmTblcombi<-rlmTblcombi[!duplicated(rlmTblcombi$treatment), ] ###remove duplicated rows from table rlmTblcombi
rlmTblcombi$pvalue<-(2-2*pnorm(abs(rlmTblcombi$tvalue))) ###calculate p-values
rlmTblcombi$pVal_fdr<-p.adjust(rlmTblcombi$pvalue,method="fdr") ###apply fdr correction to p-values
rlmTblcombi$pVal_fdr<-as.numeric(rlmTblcombi$pVal_fdr)

#Calculate hits of robust linear model

#Preparations
CombiTbl<-data.frame(treatment=character())  #Create new data frame with all the gene combinations
CombiTblall<-data.frame(treatment=character())
for (i in 1:nrow(geneTbl2)){
    for (j in 1:nrow(geneTbl2)){
        CombiTbl<-data.frame(paste(geneTbl2$gene[i],"__",geneTbl2$gene[j]))
        CombiTblall<-rbind(CombiTblall,CombiTbl)
    }
}
names(CombiTblall)<-c("treatment")
for (i in 1:nrow(CombiTblall)){
    geneName<-unlist(strsplit(as.character(CombiTblall$treatment[i])," __ "))  #Separate gene names in table CombiTblall
    CombiTblall$Gene1[i]=geneName[1] ### create new col in Tblall with gene1
    CombiTblall$Gene2[i]=geneName[2] ### create new col in Tblall with gene2
    CombiTblall$treatment[i]<-paste(CombiTblall$Gene1[i],"__",CombiTblall$Gene2[i])
}

#Calculation of hits
rlmTblfdrsign1star<-subset(rlmTblcombi,rlmTblcombi$pVal_fdr<0.05)
rlmTblfdrsign2stars<-subset(rlmTblcombi,rlmTblcombi$pVal_fdr<0.01)
rlmTblfdrsign3stars<-subset(rlmTblcombi,rlmTblcombi$pVal_fdr<0.001)
```

```
rlmTblfdrsign4stars<-subset(rlmTblcombi,rlmTblcombi$pVal_fdr<0.0001) ###Create four new tables that contain the "hits"
###Define significance as follows: * =pVal_fdr<0.05, **= pVal_fdr<0.01, *** =pVal_fdr<0.001, ###****=pVal_fdr<0.0001
rlmTblfdrsign1star[,"robustZscore"] <- NA
rlmTblfdrsign1star[,"madRobZscore"] <- NA
for (l in 1:nrow(rlmTblfdrsign1star)){
 for (k in 1:nrow(CombiTblall)){
  if ((CombiTblall$treatment[k]==rlmTblfdrsign1star$treatment[l])){
   rlmTblfdrsign1star$robustZscore[l]<-CombiTblall$robustZscore[k]
   rlmTblfdrsign1star$madRobZscore[l]<-CombiTblall$madRobZscore[k]
  }
 }
}
###add robustZscore and madRobZscore values to hits table, from CombiTblall rlmTblfdrsign2stars[,"robustZscore"] <- NA
rlmTblfdrsign2stars[,"madRobZscore"] <- NA
for (l in 1:nrow(rlmTblfdrsign2stars)){
 for (k in 1:nrow(CombiTblall)){
  if ((CombiTblall$treatment[k]==rlmTblfdrsign2stars$treatment[l])){
   rlmTblfdrsign2stars$robustZscore[l]<-CombiTblall$robustZscore[k]
   rlmTblfdrsign2stars$madRobZscore[l]<-CombiTblall$madRobZscore[k]
  }
 }
}
rlmTblfdrsign3stars[,"robustZscore"] <- NA
rlmTblfdrsign3stars[,"madRobZscore"] <- NA
for (l in 1:nrow(rlmTblfdrsign3stars)){
 for (k in 1:nrow(CombiTblall)){
  if ((CombiTblall$treatment[k]==rlmTblfdrsign3stars$treatment[l])){
   rlmTblfdrsign3stars$robustZscore[l]<-CombiTblall$robustZscore[k]
   rlmTblfdrsign3stars$madRobZscore[l]<-CombiTblall$madRobZscore[k]
  }
 }
}

rlmTblfdrsign4stars[,"robustZscore"] <- NA
rlmTblfdrsign4stars[,"madRobZscore"] <- NA
for (l in 1:nrow(rlmTblfdrsign4stars)){
 for (k in 1:nrow(CombiTblall)){
  if ((CombiTblall$treatment[k]==rlmTblfdrsign4stars$treatment[l])){
   rlmTblfdrsign4stars$robustZscore[l]<-CombiTblall$robustZscore[k]
   rlmTblfdrsign4stars$madRobZscore[l]<-CombiTblall$madRobZscore[k]
  }
 }
}

#Save results
rsignfdrsheet1<-createSheet(wb, sheetName="rsignfdrTbl1star")
addDataFrame(x=rlmTblfdrsign1star,sheet=rsignfdrsheet1, row.names=F)
rsignfdrsheet2<-createSheet(wb, sheetName="rsignfdrTbl2stars")
addDataFrame(x=rlmTblfdrsign2stars,sheet=rsignfdrsheet2, row.names=F)
rsignfdrsheet3<-createSheet(wb, sheetName="rsignfdrTbl3stars")
addDataFrame(x=rlmTblfdrsign3stars,sheet=rsignfdrsheet3, row.names=F)
rsignfdrsheet4<-createSheet(wb, sheetName="rsignfdrTbl4stars")
addDataFrame(x=rlmTblfdrsign4stars,sheet=rsignfdrsheet4, row.names=F)
saveWorkbook(wb,"/Volumes/t2pepperkok/Anthi/ScreenAnalysis/Interactions.xlsx")
```

7.4.3.2 ALTERNATIVE APPROACH; SAMPLE-BASED NORMALIZATION

An alternative statistical method was used, to analyze the co-RNAi screen data for genetic interactions. In this approach, the effect on LDL uptake of the combined transfection with two siRNAs targeting two different genes A and B was compared to the additive effect of the two single transfections, which were considered the combination of the respective siRNA with any other siRNA. For this, a robust linear

model was fitted using R, where the additive effect of two single knockdowns A and B {($Gene_A$+any)+($Gene_B$+any)} was regressed to the respective double knockdown effect ($Gene_A$+$Gene_B$), assuming the negative control as an intercept, to derive a common effect estimate for each combination (Cordell 2002). Thus, the interaction effect w_{AB} between two genes A and B was calculated from the difference between the main effects m_A and m_B of the two single knockdowns and the effect of the double knockdown y_{AB} (w_{AB}= m_A+m_B-y_{AB}-y_0), with y_0 being the baseline value, which is given from the median effect of any other treatment. A p-value was then calculated and adjusted using the false discovery rate (fdr) method, as before. The R script for the interaction calling is shown below, with descriptions of the steps.

```
#######
# Script written by Anthi Trasta with the help of Bernd Klaus from Huber group
#Combinatorial RNAi screen with LDL-uptake assay
#Interaction calling

###The parts; Preparations, Import data, Split treatment names and Prepare for rlm fitting are the same as in previous code

###Fit robust linear model
rlmTbl<-data.frame()
for(i in 1:nrow(geneTbl)) {
    testGene1<-geneTbl$gene[i]
    for(j in 1:nrow(geneTbl)){
        testGene2<-geneTbl$gene[j]
        if (testGene1!=testGene2){
        ### create new data frame, with only those combinations from "Tblall" that contain the two testGenes
        ### combinations: Gene1__any,Gene2__any, Gene1__Gene2 and [any treatment] as control
        ### the "testTbl" is overwritten in each run of the loop
        controlTbl<-subset(Tblall,((Tblall$Metadata_gene_sorted!="INCENP__NA")&
        (Tblall$Metadata_gene_sorted!="NPC1__NA")&(Tblall$Metadata_gene_sorted!="NEG9__NPC1")))
        controlTbl$Metadata_gene_sorted<-"any"
        controlTbl$Gene1<-"any"
        controlTbl$Gene2<-"any"
        single1Tbl<-subset(Tblall,((Gene1==testGene1)|(Gene2==testGene1)))
        for(k in 1:nrow(single1Tbl)){
            single1Tbl$Gene2[k]<-"any"
            single1Tbl$Gene1[k]<-testGene1
            single1Tbl$Metadata_gene_sorted<-paste(single1Tbl$Gene1,"__",single1Tbl$Gene2)
        }
        single2Tbl<-subset(Tblall,((Gene1==testGene2)|(Gene2==testGene2)))
        for(k in 1:nrow(single2Tbl)){
            single2Tbl$Gene2[k]<-"any"
            single2Tbl$Gene1[k]<-testGene2
            single2Tbl$Metadata_gene_sorted<-paste(single2Tbl$Gene1,"__",single2Tbl$Gene2)
        }
        doublesTbl<-subset(Tblall,((Gene1==testGene2)&(Gene2==testGene1))|((Gene1==testGene1)&(Gene2==testGene2)))
        for(k in 1:nrow(doublesTbl)){
            doublesTbl$Gene1[k]<-testGene1
            doublesTbl$Gene2[k]<-testGene2
            doublesTbl$Metadata_gene_sorted<-paste(doublesTbl$Gene1,"__",doublesTbl$Gene2)
        }
        testTbl<-rbind(controlTbl,single1Tbl,single2Tbl,doublesTbl)
        combi<-((testTbl$Gene1==testGene1)&(testTbl$Gene2==testGene2)) ### combi: testGene1_testGene2
        single1<-((testTbl$Gene1==testGene1)&(testTbl$Gene2=="any")|combi) ### single1: testGene1_any
        single2<-((testTbl$Gene1==testGene2)&(testTbl$Gene2=="any")|combi) ### single2: testGene2_any
        rlmfit<-rlm(robustZscore~single1+single2+combi,data=testTbl)
        rs <- summary(rlmfit)$sigma
        rlmfit.df<-as.data.frame(coef(summary(rlmfit)))
        names(rlmfit.df)[2]<-"Std.Error"
```

```
        rlmfit.df$Gene1<-paste(testGene1)
        rlmfit.df$Gene2<-paste(testGene2)
        fsheet<-createSheet(wb2, sheetName=paste(testGene1,"__",testGene2))
        addDataFrame(x=rlmfit.df,sheet=fsheet, row.names=F)
        rlmTbl<-rbind(rlmTbl,rlmfit.df)
        }
    }
}
names(rlmTbl)[3]<-"tvalue" ###name the 3rd column of rlm Tbl
rlmTbl$coefficient=rownames(rlmTbl) ###create new column in lmTbl, "coefficient", copying the column "rownames"

###The rest of the code is the same as in previous one
```

7.4.4 Plotting of Screen Hits

7.4.4.1 *Barplots*

```
### This R-code chunk continues from the previous one
rlmTblfdrsign2stars<-rlmTblfdrsign2stars[with(rlmTblfdrsign2stars,
order(-rlmTblfdrsign2stars$Value,rlmTblfdrsign2stars$treatment)),] ###order hits according to interaction effect

pdf("/Volumes/t2pepperkok/Anthi/ScreenAnalysis/barplot_interactions_2stars.pdf")   par(mar=c(8,7,2,2))
bp<-barplot(height=rlmTblfdrsign4stars$Value,
                space=c(0,1),
                main="Gene Interactions pVal(fdr)<10^-4",
                axes=F,
                ylim=c(-4,4),
               xpd=F )
axis(side=2,
     pos=1,
     las=1,
     at=c(-4,-3,-2,-1,0,1,2,3,4))
axis(side=1,
       at=bp,
       labels=F)
text(x=seq_along(rlmTblfdrsign4stars$treatment),
       par("usr")[3]-0.2,
       adj=c(0.9,2.4),
       srt=45,xpd=T,
       labels=rlmTblfdrsign4stars$treatment,
       cex=0.7)
arrows(bp,rlmTblfdrsign4stars$Value+rlmTblfdrsign4stars$Std.Error,      bp,rlmTblfdrsign4stars$Value-
       rlmTblfdrsign4stars$Std.Error,
       angle=90,
       code=3,
       length=0.05)
abline(h=c(0))
mtext("treatment",side=1,line=6) mtext("Estimated value of robust linear model",side=2,line=5)

dev.off()
```

7.4.4.2 Correlation Heatmap

```
### This R-code chunk continues from the previous one
CombiTblalln<-data.frame(grep("NEG9",CombiTblall$treatment,value=T,invert=T))
CombiTblalln<-subset(CombiTblall,(CombiTblall$Gene1!="NEG9")&(CombiTblall$Gene2!="NEG9"))

pdf("/Volumes/t2pepperkok/Anthi/ScreenAnalysis/CorrelationHeatmap.pdf")
rgb.palette <- colorRampPalette(c("green", "white","magenta"), space = "rgb")
par(mar=c(8,6,2,2))

#Build the horizontal and vertical axis information
hor <- c(geneTbl$gene)
ver <- c(geneTbl$gene)

#Build the fake correlation matrix, see matrix{base}
Nrow<-length(unique(as.factor(CombiTblalln$Gene1)))
Ncol<-length(unique(as.factor(CombiTblalln$Gene2)))
```

```
CorHeatDat<-CombiTblalln$robustZscore
cor <- matrix(data=CorHeatDat, nrow=Nrow, ncol=Ncol, dimnames = list(hor, ver))

#Build the plot
levelplot(cor,
      main="Candidate LDL-effector genes array correlation matrix",
      xlab="Gene 1",
      ylab="Gene 2",
      scales=list(x=list(rot=90)),
      col.regions=rgb.palette(150),
      cuts=100,
      at=seq(-max(abs(CombiTblalln$robustZscore)),max(abs(CombiTblalln$robustZscore)),0.1))
dev.off()
```

7.4.5 qRT-PCR ANALYSIS

qRT-PCR results were analyzed using the comparative C_T method. For this, baseline fluorescence was set in the StepOne Software, to subtract the background signal owed to the reaction medium. This baseline threshold was set to 0.1 C_T and was kept constant throughout all the qRT-PCR reactions performed, for comparison reasons. Next, the threshold cycle (C_T) value was determined, which reflects the cycle number at which the fluorescence signal generated from a reaction crosses the threshold. The difference (ΔC_T) between the C_T value of a target gene and the C_T value of a housekeeping gene (typically GAPDH) was calculated. Subsequently, the difference ($\Delta\Delta C_T$) between the ΔC_T from a treated well and a non-treated well was calculated and the fold change corresponded to $2^{-\Delta\Delta CT}$. The statistical significance of the treatment fold change compared to the untreated was calculated by using a two-tailed Student's t-test in Microsoft Excel (TTEST (array1, array2, 2, 3)).

7.4.6 WESTERN BLOTTING ANALYSIS

Scanned western blots were analyzed with ImageJ to quantify protein band intensities. For the quantifications, films were selected where bands were not saturated. As a first step in the protein quantification, the intensity of bands was normalized to background intensity, and then the intensity of protein bands of interest was normalized to that of a-tubulin bands on the same gel.

7.4.7 PROTEIN/ mRNA INTERACTION ANALYSIS

To test for interactions at the level of LDLR or SREBF1/2 mRNA, or LDLR protein, the results of qPCR or western blotting experiments, respectively were used. For qPCR, results derived from three to four biological replicas, and from three replicas for

western blots. The exact same experimental conditions (48h knockdown, cell starvation) were used, as before performing the LDL-uptake assay. Specifically, for qPCR experiments, the ΔCt values (see section 7.4.5) were used to fit a linear model, whereas for Western blot experiments, the ratio LDLR/a-tubulin (see 7.4.6), without normalization was used. The effect of the combined transfection with two siRNAs targeting two different genes A and B was compared to the additive effect of the two single transfections. In single transfections the respective siRNAs were co-transfected together with the negative control siRNA. For this, a robust linear model was fitted using R, where the additive effect of two single knockdowns A and B {($Gene_A$+control)+($Gene_B$+control)} was regressed to the respective double knockdown effect ($Gene_A$+$Gene_B$), assuming the negative control as an intercept, to derive a common effect estimate for each combination (Cordell 2002). A p-value was then calculated from the t-value of the linear regression model using the following equation; p_{Val}=2-2*pnorm(abs(t_{Val})). To correct for multiple comparison, the p-value was adjusted using the false discovery rate (fdr) method (Benjamini and Hochberg 1995), and the significance threshold for gene interactions was set at $pValue_{fdr}<10^{-1}$. The R script for the interaction calling is shown below, with descriptions of the steps.

```
#######
# Script written by Anthi Trasta
rm(list=ls())         ### clear R-memory

#Preparations

library(xlsx)         ### load xlsx package
library(MASS)      ### load MASS package for rlm function
setwd("/Users/anthi/Desktop/Interaction_calc")

#Import data
Tbl<-read.csv("Table.csv") ### read file and save to vector Tbl

#Split treatments into two gene names
for (i in 1:nrow(Tbl)){
  geneName<-unlist(strsplit(as.character(Tbl$treatment[i]),"__"))

  Tbl$Gene1[i]=geneName[1] ### create new col in Tblall with gene1
  Tbl$Gene2[i]=geneName[2] ### create new col in Tblall with gene2

}

#Prepare for rlm fitting
###Change the two test genes every time this chunk is ran, to test for all gene pairs
testGene1<-"HAVCR1"
testGene2<-"LDLR"
rlmTbl<-data.frame()
testTbl<-subset(Tbl,(Gene1=="NEG9")|((Gene1==testGene1)&(Gene2=="NEG9"))|((Gene1==testGene2)&(Gene2=="NEG9"))|
          ((Gene1==testGene1)&(Gene2==testGene2))|((Gene1==testGene2)&(Gene2==testGene1)))
combi<-
(((testTbl$Gene1==testGene1)&(testTbl$Gene2==testGene2))|((testTbl$Gene1==testGene2)&(testTbl$Gene2==testGene1)))
```

```
single1<-((testTbl$Gene1==testGene1)&(testTbl$Gene2=="NEG9")|combi)
single2<-((testTbl$Gene1==testGene2)&(testTbl$Gene2=="NEG9")|combi)

#Fit robust  linear model
rlmfit<-rlm(DDCt~single1+single2+combi,data=testTbl)
rs <- summary(rlmfit)$sigma
rlmfit.df<-as.data.frame(coef(summary(rlmfit)))
names(rlmfit.df)[2]<-"Std.Error" #name column 2 of generated table
rlmfit.df$Gene1<-paste(testGene1)
rlmfit.df$Gene2<-paste(testGene2)
rlmTbl<-rbind(rlmTbl,rlmfit.df)
names(rlmTbl)[3]<-"tvalue" #name column 3 of generated table
rlmTbl$pvalue<-(2-2*pnorm(abs(rlmTbl$tvalue))) #calculate p-values

#Save data
wb<-createWorkbook()
sheet1<-createSheet(wb, sheetName="HAVCR1_LDLR")
addDataFrame(x=rlmTbl,sheet=sheet1, row.names=F)
saveWorkbook(wb, "/Users/anthi/Desktop/qRT-PCR/LDLR_mRNA_interaction_calc/linear_model.xlsx")
```

7.4.8 NUCLEIC ACID DESIGN

7.4.8.1 *qRT-PCR PRIMER DESIGN*

Primers for qRT-PCR were custom designed to target the genes studied and met certain requirements, according to which the parameters of the primer design tool (*NCBI Primer-Blast*) were set; First, the primers need to span an exon-exon junction to avoid amplification of genomic DNA. For this purpose, two neighboring exons were selected that were present in most protein coding transcripts of the gene, using the Ensembl genome browser. The sequences of the two exons were extracted from Ensembl and jointly pasted into NCBI Primer-Blast Tool. The range for the design of the forward primer was set within the sequence of the first exon, while the range for the design of the reverse primer was set within the sequence of the second exon. Second, the PCR product size was set between 60 and 150 bp for efficient amplification. Third, the primer melting temperature was set between 57°C and 63°C, with the ideal T_m being 60°C, and the maximum T_m between the two primers targeting the same gene was set as 3°C. Fourth, the primer GC content was set around 50-60% to ensure maximum product stability. Last, the primer length was set between 18 and 23 nucleotides. In order to examine the retrieved primers from NCBI Primer-Blast for specific targeting, the primers were mapped against the human reference genome (Human GRCh38) using the Ensembl BLAST/BLAT Tool. Specifically, each primer sequence was pasted into BLAST, and a search was performed against the DNA database, selecting the option “Genomic sequence”, and with the search tool BLASTN, with sensitivity for Short sequences. Primers binding to

other genomic locations with high identity and statistical significance were not selected. Next, selected primers were evaluated for secondary structure formation and dimerization probability using the OligoEvaluator Tool from Sigma-Aldrich. Primers with high probability to form secondary structures were rejected. Finally, the selected primers that passed the aforementioned quality control criteria were ordered from Sigma Aldrich.

7.4.8.2 siRNA DESIGN

7.4.8.2.1 MAPPING siRNAs TO THE REFERENCE GENOME

Most of the siRNAs used in this study were extracted from Blattmann et al. (Blattmann et al. 2013), and few were selected using Bluegecko Tool (J.K. Hériché). All siRNAs were initially mapped to the reference genome (Ensembl 66 release GRCh38) with Bluegecko. With this tool the number of targeted protein-coding transcripts was evaluated and unspecific siRNAs that target other human mRNAs were identified. Moreover, the siRNAs were evaluated for mismatches to the reference sequence of the respective target gene or for targeting transcripts not anymore considered as protein coding.

7.4.8.2.2 CELLBASE

All the plates and microscope images from the interaction screen performed in this study were uploaded and stored in Cellbase. Cellbase is an EMBL internal database, developed and maintained by S. Sauer. Thus, all the images from spotted positions across a LabTek could be easily explored and evaluated.

8 ABBREVIATIONS

CAD	Coronary Artery Disease
CDCV	Common Disease, Common Variant
cDNA	complementary DNA
CDRV	Common Disease, Rare Variant
Co-RNAi	combinatorial RNA interference
Cq	quantitation cycle
Ct, C_T	threshold cycle
Ctrl	Control
CVAS	Common Variant Association Study
CVD	Cardiovascular Disease
DAPI	4,6-Diamino-2-Phenylindole
DiI	Di-alkyl-Indocarbocyanine dye
DMEM	Dulbecco's Modified Eagle Medium
DMSO	Dimethylsulfoxide
DRAQ5	Deep Red Anthraquinone 5 EDTA
Em	Emission
ESP	Exome Sequencing Project
EVS	Exome Variant Server
Ex	Excitation
FC	Free Cholesterol
FH	Familial Hypercholesterolemia
FW	Forward
GAPDH	Glyceraldehyde 3-Phosphate Dehydrogenase
GFP	Green Fluorescent Protein
GI	Genetic Interaction
GWAS	Genome Wide Association Study
HDL	High-Density Lipoprotein
HDL-C	HDL-cholesterol
HPCD	Hydroxypropylcyclodextrin

IM	Imaging Medium
LD	Linkage Disequilibrium
LDL	Low-Density Lipoprotein
LDL-C	LDL-cholesterol
LDLR	Low-Density Lipoprotein Receptor
MAD	Median Absolute Deviation
MAF	Minor Allele Frequency
MI	Myocardial Infarction
MOPS	(3-(N-morpholino) propanesulfonic acid
mRNA	messenger RNA
OR	Odds Ratio
PBS	Phosphate Buffer Saline
PBS-T	Phosphate Buffer Saline-Tween
PCR	Polymerase Chain Reaction
PFA	Paraformaldehyde
PVDF	Polyvinylidene Fluoride
RNAi	RNA-mediated interference
RNA-Seq	RNA Sequencing
RT	Room Temperature
RT-qPCR	quantitative Real-Time Polymerase Chain Reaction
RV	Reverse
RVAS	Rare Variant Association Study
SDS-PAGE	Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis
siRNA	small interfering RNA
SNP	Single Nucleotide Polymorphism
TG	Triglycerides
TC	Total Cholesterol
ΔC_T	threshold cycle difference

9 APPENDIX

9.1 SUPPLEMENTARY TABLES

Table S1. Exome-Chip results from 2/12/2013 (Peloso et al. 2014)
Shown in this table are the number of individuals genotyped for each of four lipid traits (LDL, HDL, TG and TC), and the number of variants identified at a genome-wide significance level to be associated with these traits, as well as the number of the corresponding genes. Moreover, the number of genes that were also tested in Blattmann et al. ("of these in GWAS-RNAi"), and the percentage of overlap between genes identified in Exome Chip and the ones from Blattmann et al. ("% overlap") are shown.

	LDL	**HDL**	**TG**	**TC**
No. of individuals genotyped	73.652	75.995	75.273	76.948
No. of variants with p<10-7	159	242	182	207
No. genes with variants p<10-7	61	87	56	68
of these in GWAS-RNAi	28	28	25	35
% overlap	0,26	0,18	0,23	0,27
No. genes with missense p<10-7	33	55	29	37
of these in GWAS-RNAi	13	18	13	22
% overlap	0,28	0,23	0,33	0,44

Table S2. Exome-Chip results from 2/12/2013 (Peloso et al. 2014) **and comparison of data to Blattmann et al., by Heiko Runz.**

Juxtaposition of the Exome Chip results with the Blattmann et al results. In the left part of the table are shown are the number of variants (total and missense) from the Exome Chip associated at a genome-wide significance level with each of four lipid traits (LDL, HDL, TG and TC). In the right part of the table are shown the number of siRNAs tested per gene in the Blattmann et al. study ("siRNAs"), and the number of siRNAs that had an effect on LDL uptake ("LDL") or free cholesterol ("FC"). Moreover, it is shown if the gene was tested in a secondary assay ("sec.assay?"), y= tested, n=not tested, if the filipin assay results were validated in Huh7 cells ("FC val."), or in if it had an effect on LDLR mRNA ("LDLR"), if the overexpression of GFP-tagged protein of the gene ("GFP-"), or of its cDNA had an effect ("cDNA"). SiRNAs that showed an effect below or above thresholds are indicated with blue (decreased) or red (increased).

Gene	Exome Chip (2-12-13; 75K individuals)								Blattmann et al., 2013							
	LDL		HDL		TG		TC		RNAi effector?			Secondary assay effector?			cDNA effector?	
	total	miss.	total	miss.	total	.miss.	total	miss.	siRNAs	LDL	FC	sec. assay?	FC val.	LDLR	GFP-	cDNA
APOE	2	1	2	1	2	1	2	1	5	1	3	y	2	NA	NA	NA
BAZ1B	0	0	2	0	3	0	0	0	6	0	3	y	2	0	y	2
BCAM	2	2	0	0	1	1	2	2	6	0	3	n	NA	NA	NA	NA
BCL7B	0	0	1	0	1	0	0	0	5	1	4	y	0	-2	NA	NA
CBLC	1	1	0	0	0	0	1	1	4	2	2	y	0	NA	NA	NA
CELSR2	4	1	2	0	0	0	4	1	6	0	1	n	NA	NA	NA	NA
HAVCR1	0	0	0	0	0	0	1	1	6	0	4	y	2	NA	y	2
LDLR	3	0	0	0	0	0	3	0	7	-4	-2	y	-2	-2	y	2
LPL	0	0	8	3	8	3	0	0	6	-1	1	y	NA	0	y	0
MLXIPL	0	0	2	2	2	2	0	0	5	0	2	y	2	2	y	0
MYBPHL	2	2	0	0	0	0	1	1	5	1	3	y	0	NA	y	2
NCAN	1	1	0	0	1	1	1	1	5	1	4	y	2	NA	y	2
PCSK9	3	3	0	0	0	0	3	3	3	0	3	n	NA	NA	y	NA
PVRL2	1	0	0	0	0	0	1	0	4	1	1	n	NA	NA	NA	NA
SIK3	0	0	1	0	3	0	2	0	5	0	1	y	NA	0	y	2
TM6SF2	1	1	0	0	1	1	1	1	5	-1	3	y	2	2	y	0
TMEM57	1	0	0	0	0	0	1	0	5	1	3	y	2	0	y	2
TOMM40	2	0	2	0	2	0	2	0	5	-1	3	y	-2	NA	y	0
APOB	9	8	4	3	5	4	7	6	3	0	3	y	0		NA	NA
HMGCR	3	0	0	0	0	0	3	0	3	0	1	y	NA	0	NA	NA
ZNF259	1	0	3	1	3	1	3	1	4	0	1	n	NA	NA	NA	NA
CXCL12	0	0	0	0	0	0	0	0	5	1	3	y	0	1	y	-1
SORT1	0	0	0	0	0	0	0	0	4	-1	-1	y	-1	0	y	1
FAM174A	0	0	0	0	0	0	0	0	5	-1	2	y	0	-2	y	2
PAFAH1B1	0	0	0	0	0	0	0	0	5	-3	-2	y	-2	-2	y	0
WDR12	0	0	0	0	0	0	0	0	0	0	-2	y	0	-2	y	2
SEZ6L	0	0	0	0	0	0	0	0	5	-1	-1	y	-1	2	y	1
LDLRAP1	0	0	0	0	0	0	0	0	5	0	3	y	1	4	NA	NA
MYLIP	1	1	0	0	0	0	0	0	NA	NA	NA	NA	NA	NA	y	NA
PAFAH1B2	0	0	1	1	3	2	2	1	NA	NA	NA	NA	NA	NA	NA	NA

Table S3. Overview of genes analyzed and the GWAS that showed association to blood lipid levels and/or CAD/MI.

Locus	1p13.3	1p32.3	1p36.11	2p24-p23	2q33	5q13.3-q14	5q21	5q23	6p22.3-23	7q11. 23	8p21.3	10q11.1	11q23.3	17q13.3	19p12	19p13. 2	19q13. 2	22q12.1
Genes	CELSR2/ MYBPHL/ SORT1	PCSK9	LDLRAP1/ TMEM57	APOB	WDR12	HMGCR	FAM174A	HAVCR1	MYLIP	BAZ1B/ BCL7B/ MLXIPL	LPL	CXCL12	SIK3/ ZNF259/ PAFAH1B2	PAFAH1B1	NCAN/ TM6SF2	LDLR	APOE/ BCAM/ CBLC/ PVRL2/ TOMM40	SEZ6L
Lead trait	LDL	LDL	TC	LDL/TG	CAD/MI	TC	CAD/MI	TC	TG	TG	TG	CAD/MI	TG	HDL	TC	LDL	LDL/TG	CAD/MI
studies with association	28	21	6	29	4	13	2	8	5	28	37	9	28	4	8	24	41	4
Allen 2016																	x	
Angelakopoulou 2012				x													x	x
Asselberg et al. 2012	x	x		x		x				x	x		x		x	x		
Aulchenko et al. 2009	x		x	x		x				x	x		x		x	x	x	
Barber et al. 2010	x			x						x				x				
Below 2016	x										x		x		x			
Benn 2008						x												
Braun et al. 2012	x												x				x	
Brautbar et al. 2011													x					
Burkhardt et al. 2008						x											x	
C4D et al. 2012	x	x		x	x						x	x				x		
Calandra et al. 2011	x			x		x			x									
Chambers et al. 2011										x			x					
Chasman et al. 2008				x									x				x	
Chasman et al. 2009a										x	x							
Chasman et al. 2009b				x														
Chasman et al. 2012		x							x							x	x	
Comuzzie et al. 2012													x					
Coram et al. 2013		x		x							x					x	x	
Dichgans et al. 2013	x				x						x		x			x		
Elbers et al. 2012		x														x	x	
Willer GLGC 2013	x	x	x	x		x			x	x	x		x	x	x	x	x	
Grallert et al. 2012													x					
Guo et al. 2011										x								
Heid et al. 2008											x							
Hopewell et al. 2013																	x	
Johansen et al. 2011											x		x					
Kamatani et al. 2010										x	x							
Kathiresan et al. 2008a	x			x		x		x		x	x		x		x	x	x	
Kathiresan et al. 2008b		x		x		x											x	
Kathiresan et al. 2009	x	x			x					x	x	x				x		
Katoh et al. 2004													x					
Keller et al. 2013											x							
Kettunen et al. 2012		x								x			x				x	
Kim et al. 2011										x	x		x				x	
Ko et al. 2014											x		x					
Kooner et al. 2008										x	x		x					
Kraja et al. 2011											x		x				x	
Kristiansson et al. 2012										x							x	
Kurano 2016											x		x				x	
Lee 2013												x						
Lettre et al. 2011	x	x		x		x		x			x	x		x		x		x
Lu et al. 2010											x					x	x	
Makela et al. 2014				x														
Mehta et al. 2011												x						
Middelberg et al. 2011				x							x		x			x	x	
MIGC 2009		x																
Musunuru et al. 2010	x																	
Musunuru et al. 2012	x	x		x		x				x	x		x			x	x	
Nakayama et al. 2011	x									x								
Nikpay et al. 2015	x	x		x								x				x	x	
Pan et al. 2009										x								
Parihar 2014				x						x	x							
Rasmussen et al. 2012																	x	
Reilly et al. 2013	x										x		x				x	
Reynolds et al. 2010										x								
Ridker et al. 2009											x		x				x	
Ronald et al. 2009																	x	
Sabatti ei al. 2009	x			x										x		x	x	
Saleheen et al. 2010													x					
Samani et al. 2007	x						x				x	x						x
Sandhu et al. 2008	x			x													x	
Saxena et al. 2007				x							x						x	
Schunkert et al. 2011	x	x			x							x	x			x		
Service et al. 2014											x							
Seshadri et al. 2010																	x	
Shen et al. 2010				x														
Smith et al. 2010																	x	
Surakka 2015		x	x	x				x	x		x				x	x	x	
Song et al. 2013	x																	
Talmud et al. 2009										x			x					
Tan et al. 2012											x		x					
Teslovich et al. 2010	x	x	x	x		x		x	x	x	x		x			x	x	
Tukainen et al. 2012	x	x	x	x		x					x					x	x	
Uebi et al. 2012													x					
Vrablik et al. 2008										x								
Wakil 2016												x						
Wallace et al. 2008	x			x						x	x		x			x	x	
Wang et al. 2008										x								
Waterworth et al. 2010	x	x	x	x		x		x	x	x	x		x			x	x	
Weissglas-Volkov et al. 2013										x								
Willer et al. 2008	x	x		x				x		x	x		x		x	x	x	
WTCC 2007							x										x	x
Wu et al. 2013				x				x		x	x						x	
Zhang et al. 2011								x					x					
Zhou et al. 2011											x		x		x		x	

Table S4. The 30 genes selected for the co-RNAi screen and rationale for selection

Shown are the genes that were screened for gene-gene interactions in the co-RNAi screen, and the reason why they were selected. If the gene was a hit (+) in both studies, no other reason is provided. If the gene was a hit only in one of the two studies, the reason for selection is given in the column "other".

	Gene	Ensemble ID	Reason for selection		
			Blattmann et al. hit	Exome chip hit	Other
1	APOB	ENSG00000084674	+	+	
2	APOE	ENSG00000130203	+	+	
3	BAZ1B	ENSG00000009954	+	+	
4	BCAM	ENSG00000187244	+	+	
5	BCL7B	ENSG00000106635	+	+	
6	CBLC	ENSG00000142273	+	+	
7	CELSR2	ENSG00000143126	+	+	
8	CXCL12	ENSG00000107562	+	-	one of the strongest hits in Blattmann et al.
9	FAM174A	ENSG00000174132	+	-	
10	HAVCR1	ENSG00000113249	+	+	
11	HMGCR	ENSG00000113161	+	+	
12	LDLR	ENSG00000130164	+	+	
13	LDLRAP1	ENSG00000157978	+	-	1) required for efficient endocytosis of LDLR 2) mutations are responsible for autosomal recessive hypercholesterolemia
14	LPL	ENSG00000175445	+	+	
15	MLXIPL	ENSG00000009950	+	+	
16	MYBPHL	ENSG00000221986	+	+	
17	MYLIP	ENSG00000007944		+	degrader of LDLR
18	NCAN	ENSG00000130287	+	+	
19	PAFAH1B1	ENSG00000007168	+	-	potential significant role in cholesterol regulation, based on P. Blattmann results
20	PAFAH1B2	ENSG00000168092	-	+	forms complex together with PAFAH1B1
21	PCSK9	ENSG00000169174	+	+	
22	PVRL2	ENSG00000130202	+	+	
23	SEZ6L	ENSG00000100095	+	-	one of the strongest hits in Blattmann et al.
24	SIK3	ENSG00000160584	+	+	
25	SORT1	ENSG00000134243	+	-	1)potential role in atherosclerosis 2)one of the strongest hits in Blattmann et al.
26	TM6SF2	ENSG00000213996	+	+	
27	TMEM57	ENSG00000204178	+	+	
28	TOMM40	ENSG00000130204	+	+	
29	WDR12	ENSG00000138442	+	-	one of the strongest hits in Blattmann et al.
30	ZNF259	ENSG00000109917	+	+	

Table S5. Genes analyzed in the co-RNA screen and a priori knowledge on molecular function and association with cholesterol regulation and/or lipid homeostasis.

Genes tested in the co-RNAi screen are listed with HGNC Symbol and Ensembl Gene ID. Previously link of the genes to lipid metabolism using GO annotation (cellular lipid metabolism, cellular response to cholesterol, cholesterol, lipid, lipoprotein, triglyceride, high-density, low-density, very-low-density), or to monogenic lipid disorders (Teslovich, Musunuru, and Smith 2010) was checked. Genefunction adapted from www.genecards.org is shown.

	Gene	Ensembl ID	Linked to lipid homeostasis (GO)	Linked to cholesterol regulation	Described function
1	APOB	ENSG00000084674	+	+	apolipoprotein B, major constituent of chylomicrons, LDL and VLDL
2	APOE	ENSG00000130203	+	+	apolipoprotein E, mediates binding, internalization and catabolism of lipoprotein particles, ligand of LDLR, promotes reverse cholesterol transport
3	BAZ1B	ENSG00000009954	-	-	atypical protein kinase, transcription regulator
4	BCAM	ENSG00000187244	-	-	Lutheran blood group glycoprotein
5	BCL7B	ENSG00000106635	-	-	regulates positively apoptosis, role in Wnt signaling pathway
6	CBLC	ENSG00000142273	-	-	E3 ubiquitin ligase, regulator of EGFR signaling
7	CELSR2	ENSG00000143126	-	-	cadherin
8	CXCL12	ENSG00000107562	-	-	chemokine ligand, stimulates migration of monocytes and T-lymphocytes, protective role after MI
9	FAM174A	ENSG00000174132	-	-	nothing known
10	HAVCR1	ENSG00000113249	-	-	membrane receptor for HAV and TIMD4, involved in allergy, kidney injury and repair
11	HMGCR	ENSG00000113161	+	+	rate-limiting enzyme for cholesterol synthesis
12	LDLR	ENSG00000130164	+	+	binds LDL and transports it into cells by endocytosis
13	LDLRAP1	ENSG00000157978	+	+	clathrin-associated adapter protein required for efficient endocytosis of LDLR in polarized cells
14	LPL	ENSG00000175445	+	+	lipoprotein lipase, hydrolyses TGs of chylomicrons and VLDL
15	MLXIPL	ENSG00000009950	+	-	transcriptional repressor
16	MYBPHL	ENSG00000221986	-	-	nothing known
17	MYLIP	ENSG00000007944	+	+	E3 ubiquitin ligase, induces proteasomal degradation of LDLR, VLDLR, and LRP8
18	NCAN	ENSG00000130287	-	-	may regulate neuronal adhesion and neurite growth
19	PAFAH1B1	ENSG00000007168	+	-	activates Rho GTPases, required for actin polymerization, may enhance dynein-mediated microtubule sliding
20	PAFAH1B2	ENSG00000168092	-	-	Inactivates platelet-activating factor
21	PCSK9	ENSG00000169174	+	+	binds to LDLR,VLDLR,LRP1 and LRP8 and promotes their degradation in lysosomes
22	PVRL2	ENSG00000130202	-	-	receptor for herpes virus
23	SEZ6L	ENSG00000100095	-	-	may contribute to ER functions in neurons
24	SIK3	ENSG00000160584	-	-	nothing known
25	SORT1	ENSG00000134243	+	+	sorting receptor in Golgi, transports proteins from Golgi to lysosomes
26	TM6SF2	ENSG00000213996	-	-	regulates liver fat metabolism, influencing TG secretion and hepatic lipid droplet content
27	TMEM57	ENSG00000204178	-	-	nothing known
28	TOMM40	ENSG00000130204	-	-	channel-forming protein essential for import of protein precursors into mitochondria
29	WDR12	ENSG00000138442	-	-	component of the PeBoW complex, which is required for maturation of 28S and 5.8S rRNAs and formation of 60S ribosome
30	ZNF259	ENSG00000109917	-	-	signaling molecule, induces neuron differentiation and stimulates axonal growth and formation of growth cone

Table S6. Selection of siRNAs for the co-RNAi screen (Adapted from Blattmann et al. 2013)
Genes tested in the co-RNAi screen are listed with HGNC Symbol and Ensembl Gene ID. The siRNA ID, and the new siRNA ID, if it has been changed are shown, as well as Bluegecko ID and number of targeted transcripts. The effect on LDL uptake ("DiI_TotInt") and Filipin assay ("Fp_TotInt") in the two screens of Blattmann et al., and the total effect above threshold are shown. Rationale for siRNA selection is provided for the selected siRNA, which is in bold script.
*siRNA shows one or two mismatch(es) against intended gene
**siRNA targets multiple genes because of overlapping transcripts or homology and was exluded from further analysis
***gene targeted is not an established protein-coding gene

Gene information			siRNA information				first screen of Blattmann et al. (GWAS1)		second screen of Blattmann et al. (GWAS2)		effect above threshold		Rationale for siRNA selection
	Gene Name	ENSG	siRNA ID	new siRNA ID	Bluegecko ID	targeted transcripts	DiI_TotInt	Fp_TotInt	DiI_TotInt	Fp_TotInt	DiI	Fp	
1	APOB	ENSG00000084674	s1475		MCO_0062883	1/2	0.16	1.44			0	0	
			s1476		MCO_0062852	1/2	0.18	**3.36**			0	**1**	strongest effect on Fp
			s1477		MCO_0062817	1/2	0.35	2.21			0	1	
2	APOE	ENSG00000130203	s1495		MCO_0062875	4/4	0.1	2.03	0.45	0.71	0	1	
			s1496		MCO_0062909	1/4	0.65	2.23	**1.07**	1.58	**1**	1	effect on DiI
			s194291		MCO_0062832	1/4	0.17	0.16			0	0	
			s445491		-	?			0.68	0.3	0	0	
			s445492		MCO_0067615	1/4			0.58	0.6	0	0	
3	BAZ1B	ENSG00000009954	s17208		MCO_0060780	2/2	-0.42	-0.76			0	0	
			s17209		MCO_0061566	2/2	0.13	**3.15**	0.65	0.52	0	**1**	strongest effect on Fp
			s17210		MCO_0063016	2/2	-0.2	0.88			0	0	
			s445497		MCO_0067585	2/2			0.91	1.7	0	1	
			s445498		MCO_0067546	2/2			0.69	0.16	0	0	
			s445499		MCO_0067539	2/2			0.67	1.39	0	1	
4	BCAM	ENSG00000187244	s445500		MCO_0067670	1/3			0.06	0.54	0	0	
			s445501		MCO_0067651	1/3			0.78	0.05	0	0	
			s8336		MCO_0063049	3/3	0.03	0.19			0	0	
			s8337		MCO_0063024	3/3	0.28	-0.88			0	0	
			s8338		MCO_0062977	3/3	0.66	**1.53**	1	0.91	0	0	strongest score on Fp
5	BCL7B	ENSG00000106635	s17733		MCO_0063008	2/2**	0.13	1.86	0.42	0.48	0	1	
			s17734		MCO_0062911	2/2	0.56	0.18			0	0	
			s228480		MCO_0063061	2/2	0.4	4.16	**1.1**	1.66	**1**	1	effect on DiI
			s445504		MCO_0067676	2/2			-0.24	-0.1	0	0	
			s445505		MCO_0067579	2/2			0.3	1.89	0	1	
6	CBLC	ENSG00000142273	**s24223**		MCO_0061739	2/2			**1.15**	**1.13**	**1**	**1**	strongest effect on DiI+ effect on Fp
			s24224		MCO_0061772	2/2			1.15	0.2	1	0	
			s445508		MCO_0067516	2/2			0.28	2.65	0	1	
			s445509		MCO_0067473	2/2			0.12	-0.38	0	0	
7	CELSR2	ENSG00000143126	s445512		MCO_0067521	1/1			0.5	0.89	0	0	
			s445513		MCO_0067513	1/1			0.12	-0.89	0	0	
			s4525		MCO_0062967	1/1	0.41	-0.02			0	0	
			s4526		MCO_0062980	1/1**	0.66	0.78	0.87	**1.7**	0	**1**	effect on Fp
			s4527		MCO_0062952	1/1	0.1	1.09	0.61	-0.24	0	0	
8	CXCL12	ENSG00000107562	s12644		MCO_0063000	6/6	0.49	3.35	0.31	1.31	0	1	
			s12645		MCO_0062869	5/6	-0.11	0.29			0	0	
			s226988		MCO_0062979	5/6**	0.8	2.21	0.68	0.67	0	1	
			s445518		MCO_0067610	2/6			**1.34**	0.64	**1**	0	effect on DiI
			s445519		MCO_0067544	2/6			-0.21	-1.98	0	-1	
9	FAM174A	ENSG00000174132	s226341		MCO_0062830	1/1	-0.12	1.5			0	0	
			s226342		MCO_0062895	1/1	0.2	-1.6	0.71	-0.69	0	-1	
			s445532		MCO_0067632	1/1			**-1.69**	1.57	**-1**	1	effect on DiI
			s445533		MCO_0067578	1/1			0.44	1.21	0	1	
			s51157		MCO_0059172	1/1	0.59	1.5	0.57	0.21	0	0	
10	HAVCR1	ENSG00000113249	**s230290**		MCO_0063002	6/6	0.37	3.7	0.33	**3.49**	0	**1**	strongest effect on Fp
			s230291		MCO_0062934	3/6	0.44	1.75	0.12	1.15	0	1	
			s25632		MCO_0063054	6/6	0.26	-0.15			0	0	
			s445544		MCO_0067639	5/6			0.41	1.04	0	0	
			s445545		MCO_0067620	5/6			0.22	0.27	0	0	
11	HMGCR	ENSG00000113161	s141		MCO_0063040	5/7	0.62	0.86			0	0	
			s142		MCO_0062915	3/7	0.40	0.81			0	0	
			s143		MCO_0062814	3/7	0.67	**1.93**			0	**1**	effect on Fp
12	LDLR	ENSG00000130164	**s224006**		MCO_0062803	8/9	**-1.47**	-4.11	-1.12	-1.34	**-1**	-1	strongest effect on DiI
			s224007		MCO_0067566	8/9			-1.13	-1.83	-1	-1	
			s224008		MCO_0062858	8/9*	0.22	0.94			0	0	
			s237197		MCO_0062790	9/9	-0.32	-0.1			0	0	
			s445552	s445551	MCO_0067650	7/9			0.59	-0.51	0	0	
			s445553	s445550	MCO_0067662	7/9			0.56	1.08	0	1	
			s6		MCO_0062944	7/9	-0.81	-0.78			-1	0	
13	LDLRAP1	ENSG00000157978	s25118		MCO_0063012	1/1	-0.15	0.77			0	0	
			s25119		MCO_0062955	1/1	0.23	1.13	0.29	1.08	0	1	
			s25120		MCO_0062881	1/1	0.36	**2.57**	0.49	0.53	0	**1**	strongest effect on Fp
			s445550	s445553	MCO_0067563	1/1			0.02	-0.28	0	0	
			s445551	s445552	MCO_0067624	1/1			-0.11	-0.72	0	0	

14	LPL	ENSG00000175445	s445556		MCO_0067630	3/4			0.42	0.08	0	0	
			s445557		MCO_0067531	3/4			0.89	1.22	0	1	
			s445558		MCO_0067471	1/4			1.01	0.13	0	0	
			s701		MCO_0062983	4/4	0.08	-0.13			0	0	
			s702		MCO_0062910	1/4	**-0.86**	-0.81	-0.09	0.33	**-1**	0	effect on DiI
			s703		MCO_0062882	2/4	0.17	1.02			0	0	
15	MLXIPL	ENSG00000009950	s27386		MCO_0062920	7/8**	0.43	2.8	0.89	0.5	0	1	rejected
			s27387		MCO_0062796	7/8	0.02	**1.99**	0.62	0.94	0	**1**	second strongest effect on Fp
			s27388		MCO_0062931	6/8	-0.2	1.69			0	0	
			s445563		MCO_0067533	6/8			0.11	-0.12	0	0	
			s445564		MCO_0067505	6/8			0.99	0.17	0	0	
16	MYBPHL	ENSG00000221986	s445565		MCO_0067649	1/1			0.26	0.14	0	0	
			s445566		MCO_0067638	1/1			0.91	0.92	0	0	
			s50998		MCO_0062849	1/1	0.35	3.74	0.25	1.68	0	1	
			s50999		MCO_0062788	1/1	0.46	3.13	**1.06**	0.55	**1**	1	effect on DiI
			s51000		MCO_0062878	1/1**	0.19	0.89			0	0	
17	NCAN	ENSG00000130287	**s3648**		MCO_0062916	2/4	0.64	3.15	**1.09**	1.08	**1**	1	effect on DiI
			s3649		MCO_0063025	2/4	0.15	0.93			0	0	
			s3650		MCO_0062843	1/4	-0.35	4.74	0.28	0.27	0	1	
			s445567	s445568	MCO_0067483	1/4			0.74	-1.1	0	-1	
			s445568		-				-0.01	1.87	0	1	
18	PAFAH1B1	ENSG00000007168	s445573		MCO_0067597	2/5			-0.44	0.86	0	0	
			s445574		MCO_0067515	2/5			**-1.2**	-0.52	**-1**	0	strongest effect on DiI
			s9996		MCO_0054981	3/5	-0.2	-3.57	-0.26	0.05	0	-1	
			s9997		MCO_0054982	4/5	-0.26	-1.35	0	-0.9	0	-1	
			s9998		MCO_0061379	4/5	-0.61	-0.28	-0.59	0.22	-1	0	
19	PCSK9	ENSG00000169174	**s48694**		MCO_0062856	1/3	0.16	**2.74**			0	**1**	strongest effect on Fp
			s48695		MCO_0062860	1/3	0.57	2.74			0	1	
			s48696		MCO_0062996	3/3	-0.22	2.22			0	1	
20	PVRL2	ENSG00000130202	s11606		MCO_0067547	3/5**			0.35	1.24	0	1	
			s11607		MCO_0067518	3/5			0.26	0.49	0	0	
			s445581		MCO_0067527	1/5			0.27	0.07	0	0	
			s445582		MCO_0067511	1/5			**1.17**	-0.23	**1**	0	effect on DiI
21	SEZ6L	ENSG00000100095	s225966		MCO_0062899	7/8	0.5	1.81	0.47	0.21	0	1	
			s24051		MCO_0063022	7/8	0.41	1.44			0	0	
			s24052		MCO_0062976	6/8	-0.49	-1.8	**-1.03**	-0.34	**-1**	-1	effect on DiI
			s445592		MCO_0067507	7/8			0.43	-1.57	0	-1	
			s445593		MCO_0067506	8/8*			0.02	-0.72	0	0	
22	SIK3	ENSG00000160584	s225956		MCO_0063064	4/5	0.43	0.97			0	0	
			s225957		MCO_0063023	4/5	0.58	0.54	0.41	0.3	0	0	
			s225958		MCO_0062890	5/5	0.7	1.08	0.79	-0.33	0	0	
			s445594		MCO_0067596	5/5			0.52	0.99	0	0	
			s445595		MCO_0067522	5/5			0.67	**1.55**	0	**1**	effect on Fp
23	SORT1	ENSG00000134243	**s224557**		MCO_0063053	2/2	**-0.56**	-2.81	-0.1	-0.38	**-1**	-1	effect on DiI
			s224558		MCO_0062806	2/2	0.29	1.37	0.49	0.67	0	0	
			s445600		MCO_0067608	2/2			0.25	0.85	0	0	
			s445601		MCO_0067591	1/2			0.52	0.5	0	0	
24	TM6SF2	ENSG00000213996	**s28703**		MCO_0062906	1/1	**-1.07**	-1.23	-0.15	0.92	**-1**	-1	effect on DiI
			s28704		MCO_0062810	1/1	0.37	3.71	0.3	2.35	0	1	
			s28705		MCO_0062958	1/1	-0.19	0.8			0	0	
			s445610		MCO_0067588	1/1			0.13	0.63	0	0	
			s445611		MCO_0067465	1/1			0.41	1.18	0	1	
25	TMEM57	ENSG00000204178	**s30473**		MCO_0062918	3/3	**0.81**	2.6	0.4	1.19	**1**	1	effect on DiI
			s30474		MCO_0062877	1/3	0.35	0.27			0	0	
			s30475		MCO_0062819	3/3	0.63	2.06	0.66	0.95	0	1	
			s445612		MCO_0067625	3/3			-0.24	-3.03	0	-1	
			s445613		MCO_0067508	3/3			0.46	0.81	0	0	
26	TOMM40	ENSG00000130204	s20448		MCO_0062829	4/6	0.33	1.07			0	0	
			s20449		MCO_0062926	4/6	-0.42	-1.63	**-1.25**	-2.16	**-1**	-1	effect on DiI
			s20450		MCO_0062835	4/6	0.62	2.51	0.77	2.17	0	1	
			s445614		MCO_0067652	4/6			-0.15	1.66	0	1	
			s445615		MCO_0067486	4/6			0.53	-0.15	0	0	
27	WDR12	ENSG00000138442	s31440		MCO_0060407	1/1	-0.09	-1.08	0.18	0.17	0	0	
			s31441		MCO_0060426	1/1	0.28	-1.65	0.5	**-1.85**	0	**-1**	effect on Fp
			s31442		MCO_0063042	1/1	0.74	-0.99			0	0	
			s445627		MCO_0067592	1/1			0.42	0.24	0	0	
			s445628		MCO_0067497	1/1			0.96	0.28	0	0	
28	ZNF259	ENSG00000109917	s16972		MCO_0067555	3/3			0.16	0.35	0	0	
			s16973		MCO_0067619	3/3**			0.44	-0.46	0	0	
			s445635		MCO_0067478	3/3			0.17	0.8	0	0	
			s445636		MCO_0067467	3/3			1.04	**1.55**	0	**1**	effect on Fp
29	MYLIP	ENSG00000007944	**s26522**		MCO_0061841	2/2							targets all transcripts
			s26523		MCO_0061848	2/2							
			s26521		MCO_0063512	1/2							
30	PAFAH1B2	ENSG00000168092	s9999		MCO_0067567	4/5***							
			s10001		MCO_0067590	5/5							targets all transcripts
			s10000		MCO_0067623	2/5							
			sPafah1b2_1		MCO_0078010	2/5							
			sPafah1b2_2		MCP_0078011	3/5							

Table S7. Amounts of siRNA used for "single" knockdowns in the validation experiments and gene knockdown efficiency, measured with qRT-PCR.

siRNA treatment	gene kd	pmol of each siRNA	% kd efficiency
HAVCR1 + control	HAVCR1	30	n.d.
		15	n.d.
LDLR + control	LDLR	30	76.8
		7.5	58.6
LDLRAP1 + control	LDLRAP1	30	92.9
MLXIPL + control	MLXIPL	30	35.6
		7.5	44.2
BCAM + control	BCAM	30	95.7
SEZ6L + control	SEZ6L	30	n.d.
		15	n.d.
SORT1 + control	SORT1	30	93.9
		7.5	91.7
TOMM40 + control	TOMM40	7.5	88.6
CXCL12 + control	CXCL12	30	n.d.
PAFAH1B1 + control	PAFAH1B1	7.5	95.3
NCAN + control	NCAN	30	n.d.
SIK3 + control	SIK3	30	78.9
MYBPHL + control	MYBPHL	30	n.d.
PCSK9 + control	PCSK9	15	85.1
TMEM57 + control	TMEM57	30	44.8
CELSR2 + control	CELSR2	30	85
LPL + control	LPL	30	96.7
APOB + control	APOB	30	n.d.
HMGCR + control	HMGCR	30	67.4

Table S8. Amounts of siRNA used for "double" knockdowns in the validation experiments and gene knockdown efficiency, measured with qRT-PCR.

siRNA treatment	gene kd	pmol of each siRNA	% kd efficiency
HAVCR1 + LDLR	HAVCR1	15	n.d.
	LDLR	7.5	58.2
HAVCR1 + LDLRAP1	HAVCR1	15	n.d.
	LDLRAP1	30	91.7
HAVCR1 + BCAM	HAVCR1	15	n.d.
	BCAM	30	94.7
HAVCR1 + MLXIPL	HAVCR1	15	n.d.
	MLXIPL	7.5	18.7
HAVCR1 + SEZ6L	HAVCR1	30	n.d.
	SEZ6L	30	n.d.
HAVCR1 + SORT1	HAVCR1	30	n.d.
	SORT1	30	93.2
BCAM + LDLRAP1	BCAM	30	94.3
	LDLRAP1	30	89.9
LDLR + MLXIPL	LDLR	30	83.4
	MLXIPL	30	39.3
LDLR + LDLRAP1	LDLR	30	75.6
	LDLRAP1	30	87.2
LDLRAP1 + SORT1	LDLRAP1	30	89.8
	SORT1	30	91.9
MLXIPL + TOMM40	MLXIPL	30	20.1
	TOMM40	7.5	90.4
CXCL12 + PAFAH1B1	CXCL12	30	n.d.
	PAFAH1B1	7.5	90.2
SORT1 + TOMM40	SORT1	7.5	52.7
	TOMM40	7.5	90.8
NCAN + TOMM40	NCAN	30	n.d.
	TOMM40	7.5	91.7
NCAN + SEZ6L	NCAN	30	n.d.
	SEZ6L	15	n.d.
NCAN + LDLR	NCAN	30	n.d.
	LDLR	7.5	72.7
PAFAH1B1 + SIK3	PAFAH1B1	7.5	77.1
	SIK3	30	77.9
MYBPHL + SIK3	MYBPHL	30	n.d.
	SIK3	30	78.6
PCSK9 + TMEM57	PCSK9	30	74.7
	TMEM57	30	51.9
CELSR2 + LPL	CELSR2	30	86.7
	LPL	30	92.5
APOB + HMGCR	APOB	30	n.d.
	HMGCR	30	74.9

Table S9. siRNA quantities used for the validation of the co-RNAi screen hits

Shown are the siRNA quantities used in the liquid-phase transfections in 6-well plates (for details see 7.1.1) for the validation of the screen hits. Because the forward transfections (in liquid-phase format) were more efficient than the reverse transfections (in the co-RNAi screen), the siRNA quantities needed to be adjusted, so that the same phenotypes could be achieved as the ones observed in the screen.

	Treatment	$siRNA_A$	$siRNA_A$ (pmol)	$siRNA_B$	$siRNA_B$ (pmol)
1	LDLR+ NCAN	LDLR	7.5	NCAN	30
2	CXCL12+ PAFAH1B1	CXCL12	30	PAFAH1B1	7.5
3	NCAN + SEZ6L	NCAN	30	SEZ6L	15
4	NCAN + TOMM40	NCAN	30	TOMM40	7.5
5	HAVCR1 + MLXIPL	HAVCR1	15	MLXIPL	7.5
6	LDLRAP1 + SORT1	LDLRAP1	30	SORT1	30
7	LDLR + LDLRAP1	LDLR	30	LDLRAP1	30
8	CELSR2 + LPL	CELSR2	15	LPL	15
9	MLXIPL + TOMM40	MLXIPL	30	TOMM40	7.5
10	LDLR + MLXIPL	LDLR	30	MLXIPL	30
11	HAVCR1+ SEZ6L	HAVCR1	30	SEZ6L	30
12	HAVCR1 + LDLRAP1	HAVCR1	15	LDLRAP1	30
13	BCAM + LDLRAP1	BCAM	30	LDLRAP1	30
14	HAVCR1 + SORT1	HAVCR1	30	SORT1	30
15	BCAM + HAVCR1	BCAM	30	HAVCR1	15
16	MYBPHL + SIK3	MYBPHL	30	SIK3	30
17	APOB + HMGCR	APOB	30	HMGCR	30
18	SORT1 + TOMM40	SORT1	7.5	TOMM40	7.5
19	PAFAH1B1+ SIK3	PAFAH1B1	7.5	SIK3	30
20	PCSK9 + TMEM57	PCSK9	15	TMEM57	15
21	HAVCR1+LDLR	HAVCR1	15	LDLR	7.5

Table S10. siRNA quantities used for the evaluation of gene interactions at the LDLR mRNA level

Shown are the quantities of siRNA used in the qRT-PCR experiments for the measurement of the effect on LDLR mRNA levels of the gene interactions identified in the screen. These quantities are essentially the exactly same quantities used in the validation experiments (see Tables S8, S9). (A-D) correspond to (A-D) of Figure 21.

A

Treatment	$siRNA_A$	$siRNA_A$ (pmol)	$siRNA_B$	$siRNA_B$ (pmol)
mock	-	-	-	-
control	NEG9	30	-	-
HAVCR1	HAVCR1	15	NEG9	15
LDLR	LDLR	7.5	NEG9	7.5
LDLRAP1	LDLRAP1	30	NEG9	30
MLXIPL	MLXIPL	7.5	NEG9	7.5
BCAM	BCAM	30	NEG9	30
NCAN	NCAN	30	NEG9	30
HAVCR1+BCAM	HAVCR1	15	BCAM	30
HAVCR1+MLXIPL	HAVCR1	15	MLXIPL	7.5
HAVCR1+LDLRAP1	HAVCR1	15	LDLRAP1	30
HAVCR1+LDLR	HAVCR1	15	LDLR	7.5
BCAM+LDLRAP1	BCAM	30	LDLRAP1	30
LDLR+NCAN	LDLR	7.5	NCAN	30

B

Treatment	$siRNA_A$	$siRNA_A$ (pmol)	$siRNA_B$	$siRNA_B$ (pmol)
mock	-	-	-	-
control	NEG9	30	-	-
SORT1	SORT1	30	NEG9	30
SEZ6L	SEZ6L	30	NEG9	30
LDLRAP1	LDLRAP1	30	NEG9	30
LDLR	LDLR	30	NEG9	30
HAVCR1	HAVCR1	30	NEG9	30
MLXIPL	MLXIPL	30	NEG9	30
TOMM40	TOMM40	7.5	NEG9	7.5
LDLR+MLXIPL	LDLR	30	MLXIPL	30
LDLR+LDLRAP1	LDLR	30	LDLRAP1	30
HAVCR1+SEZ6L	HAVCR1	30	SEZ6L	30
HAVCR1+SORT1	HAVCR1	30	SORT1	30
LDLRAP1+SORT1	LDLRAP1	30	SORT1	30
MLXIPL+TOMM40	MLXIPL	30	TOMM40	7.5

C

Treatment	$siRNA_A$	$siRNA_A$ (pmol)	$siRNA_B$	$siRNA_B$ (pmol)
mock	-	-	-	-
control	NEG9	30	-	-
SORT1	SORT1	7.5	NEG9	7.5
NCAN	NCAN	30	NEG9	30
TOMM40	TOMM40	7.5	NEG9	7.5
CXCL12	CXCL12	30	NEG9	30
PAFAH1B1	PAFAH1B1	7.5	NEG9	7.5
SIK3	SIK3	30	NEG9	30
MYBPHL	MYBPHL	30	NEG9	30
SORT1+TOMM40	SORT1	7.5	TOMM40	7.5
NCAN+TOMM40	NCAN	30	TOMM40	7.5
MYBPHL+SIK3	MYBPHL	30	SIK3	30
PAFAH1B1+SIK3	PAFAH1B1	7.5	SIK3	30
CXCL12+PAFAH1B1	CXCL12	30	PAFAH1B1	7.5

D

Treatment	$siRNA_A$	$siRNA_A$ (pmol)	$siRNA_B$	$siRNA_B$ (pmol)
mock	-	-	-	-
control	NEG9	30	-	-
SEZ6L	SEZ6L	15	NEG9	15
NCAN	NCAN	30	NEG9	30
PCSK9	PCSK9	15	NEG9	15
TMEM57	TMEM57	15	NEG9	15
CELSR2	CELSR2	15	NEG9	15
LPL	LPL	15	NEG9	15
APOB	APOB	30	NEG9	30
HMGCR	HMGCR	30	NEG9	30
NCAN+SEZ6L	NCAN	30	SEZ6L	15
PCSK9+TMEM57	PCSK9	15	TMEM57	15
CELSR2+LPL	CELSR2	15	LPL	15
APOB+HMGCR	APOB	30	HMGCR	30

Table S11. Results of qRT-PCR experiments for LDLR mRNA expression used for interaction calling.
Shown are the ΔCt values of LDLR expression, where ΔCt=Ct(LDLR)-Ct(GAPDH), for four biological replicas. The four ΔCt values were used to fit a linear regression model, where the double knockdown was compared with the additive effect of the two single knockdowns, considering NEG9 as control.

TREATMENT	siRNA quantities				mRNA expression LDLR ΔCt			
Double knockdown	**siRNA1**	**$pmol_1$**	**siRNA 2**	**$pmol_2$**	**Exp. 1**	**Exp. 2**	**Exp. 3**	**Exp. 4**
NEG9[30]	NEG9	30	NEG9	30	6.17	6.96	7.09	6.41
HAVCR1[15]+NEG9	HAVCR1	15	NEG9	30	5.94	6.15	5.88	5.75
LDLR[7.5]+NEG9	LDLR	7.5	NEG9	30	7.46	8.46	8.14	7.66
LDLRAP1[30]+NEG9	LDLRAP1	30	NEG9	30	6.98	6.98	5.55	6.34
MLXIPL[7.5]+NEG9	MLXIPL	7.5	NEG9	30	5.42	6.57	6.13	6.79
BCAM[30]+NEG9	BCAM	30	NEG9	30	7.19	7.65	6.34	6.44
HAVCR1[15]+LDLR[7.5]	HAVCR1	15	LDLR	7.5	6.74	8.57	8.07	8.17
HAVCR1[15]+LDLRAP1[30]	HAVCR1	15	LDLRAP1	30	4.98	5.66	5.66	
HAVCR1[15]+BCAM[30]	HAVCR1	15	BCAM	30	6.54	7.56	7.80	
HAVCR1[15]+MLXIPL[7.5]	HAVCR1	15	MLXIPL	7.5	5.98	5.09	5.28	6.84
BCAM[30]+LDLRAP1[30]	BCAM	30	LDLRAP1	30	6.34	6.67	6.30	6.68
SEZ6L[30]+NEG9	SEZ6L	30	NEG9	30	6.63	6.64	6.62	6.87
SORT1[30]+NEG9	SORT1	30	NEG9	30	6.18	7.12	6.68	6.53
LDLR[30]+NEG9	LDLR	30	NEG9	30	8.36	7.36	8.46	8.47
HAVCR1[30]+NEG9	HAVCR1	30	NEG9	30	5.12	5.36	5.28	6.69
MLXIPL[30]+NEG9	MLXIPL	30	NEG9	30	6.47	6.58	6.06	8.28
LDLR[30]+MLXIPL[30]	LDLR	30	MLXIPL	30	8.57	9.45	9.00	9.54
LDLR[30]+LDLRAP1[30]	LDLR	30	LDLRAP1	30	7.97	9.11	8.23	9.15
HAVCR1[30]+SEZ6L[30]	HAVCR1	30	SEZ6L	30	5.70	6.83	6.18	6.21
HAVCR1[30]+SORT1[30]	HAVCR1	30	SORT1	30	5.34	6.05	5.94	6.54
LDLRAP1[30]+SORT1[30]	LDLRAP1	30	SORT1	30	7.47		6.36	5.94
TOMM40[7.5]+NEG9	TOMM40	7.5	NEG9	30	6.45	8.15	7.05	6.46
MLXIPL[30]+TOMM40[7.5]	MLXIPL	30	TOMM40	7.5	5.58	8.90	7.18	7.02
CXCL12[30]+NEG9	CXCL12	30	NEG9	30	6.11		5.90	6.21
PAFAH1B1[7.5]+NEG9	PAFAH1B1	7.5	NEG9	30	7.08	8.20	6.62	6.35
SORT1[7.5]+NEG9	SORT1	7.5	NEG9	30	6.24	8.32	6.39	7.31
NCAN[30]+NEG9	NCAN	30	NEG9	30	5.96	6.22	6.39	6.80
CXCL12[30]+PAFAH1B1[7.5]	CXCL12	30	PAFAH1B1	7.5	6.81	7.16	6.65	7.70
SORT1[7.5]+TOMM40[7.5]	SORT1	7.5	TOMM40	7.5	6.76	8.64	7.08	7.79
NCAN[30]+TOMM40[7.5]	NCAN	30	TOMM40	7.5	7.33	7.88	7.54	8.31
SEZ6L[15]+NEG9	SEZ6L	15	NEG9	30	6.72	6.86	6.61	7.32
NCAN[30]+SEZ6L[15]	NCAN	30	SEZ6L	15	7.36	7.61	6.93	6.34
LDLR[7.5]+NCAN[30]	LDLR	7.5	NCAN	30	8.15	9.14	8.87	7.72
SIK3[30]+NEG9	SIK3	30	NEG9	30	5.94	7.15	6.50	6.49
PAFAH1B1[7.5]+SIK3[30]	PAFAH1B1	7.5	SIK3	30	7.32	6.58	6.40	5.44
MYBPHL[30]+NEG9	MYBPHL	30	NEG9	30	6.25	6.92	5.77	6.03
MYBPHL[30]+SIK3[30]	MYBPHL	30	SIK3	30	5.89	6.06	6.10	6.35
PCSK9[15]+NEG9	PCSK9	15	NEG9	30	8.31	6.14	6.15	7.54
TMEM57[15]+NEG9	TMEM57	15	NEG9	30	7.76	8.28	5.83	5.32
PCSK9[15]+TMEM57[15]	PCSK9	15	TMEM57	15	6.23	8.92	5.86	5.66
CELSR2[15]+NEG9	CELSR2	30	NEG9	30	6.82	9.16	6.18	5.59
LPL[15]+NEG9	LPL	30	NEG9	30	7.17		6.62	6.10
CELSR2[15]+LPL[15]	CELSR2	30	LPL	30	6.41		6.64	5.99
APOB[30]+NEG9	APOB	30	NEG9	30	6.57	10.78	5.82	5.29
HMGCR[30]+NEG9	HMGCR	30	NEG9	30	6.73	9.49	5.70	5.04
APOB[30]+HMGCR[30]	APOB	30	HMGCR	30	8.31	8.35	5.62	4.88

Table S12. Results of Western Blot experiments for LDLR protein used for interaction calling.
Shown are the LDLR/a-tubulin protein values without normalization to negative control siRNA, from three biological replicates used to fit a linear regression model, where the double knockdown was compared with the additive effect of the two single knockdowns, considering NEG9 as control.

blot#1	siRNA quantities				LDLR/a-tubulin		
Treatment	**siRNA1**	**pmol1**	**siRNA2**	**pmol2**	**Exp.1**	**Exp.2**	**Exp.3**
NEG9[30]	NEG9	30	NEG9	30	1.60	2.84	0.71
HAVCR1[15]	HAVCR1	15	NEG9	30	2.19	1.53	2.10
LDLR[7.5]	LDLR	7.5	NEG9	30	0.12	0.18	0.06
LDLRAP1[30]	LDLRAP1	30	NEG9	30	1.49	0.89	0.95
MLXIPL[30]	MLXIPL	30	NEG9	30	1.25	1.22	1.72
BCAM[30]	BCAM	30	NEG9	30	0.72	1.48	0.66
HAVCR1[15]+BCAM[30]	HAVCR1	15	BCAM	30	0.72	1.37	0.93
HAVCR1[15]+MLXIPL[30]	HAVCR1	15	MLXIPL	30	0.80	0.85	1.04
HAVCR1[15]+LDLRAP1[30]	HAVCR1	15	LDLRAP1	30	0.97	0.47	0.80
HAVCR1[15]+LDLR[7.5]	HAVCR1	15	LDLR	7.5	0.31	0.08	0.03

blot#2	siRNA quantities				LDLR/a-tubulin		
Treatment	**siRNA1**	**pmol1**	**siRNA2**	**pmol2**	**Exp.1**	**Exp.2**	**Exp.3**
NEG9[30]	NEG9	30	NEG9	30	1.31	1.50	1.67
HAVCR1[30]	HAVCR1	30	NEG9	30		1.15	1.62
SORT1[30]	SORT1	30	NEG9	30	1.01	0.21	0.17
SEZ6L[30]	SEZ6L	30	NEG9	30	0.80	0.43	0.42
LDLRAP1[30]	LDLRAP1	30	NEG9	30	0.90	1.20	1.34
BCAM[30]	BCAM	30	NEG9	30	0.87	0.93	0.55
LDLRAP1[30]+BCAM[30]	LDLRAP1	30	BCAM	30	0.92	0.95	1.02
LDLRAP1[30]+SORT1[30]	LDLRAP1	30	SORT1	30	1.04	0.29	0.15
HAVCR1[30]+SORT1[30]	HAVCR1	30	SORT1	30	0.67	0.78	0.30
HAVCR1[30]+SEZ6L[30]	HAVCR1	30	SEZ6L	30	0.53	1.17	0.90

blot#3	siRNA quantities				LDLR/a-tubulin		
Treatment	**siRNA1**	**pmol1**	**siRNA2**	**pmol2**	**Exp.1**	**Exp.2**	**Exp.3**
NEG9[30]	NEG9	30	NEG9	30	1.18	2.11	2.08
LDLR[30]	LDLR	30	NEG9	30	0.14	0.07	0.25
MLXIPL[30]	MLXIPL	30	NEG9	30	1.48	1.68	1.74
LDLRAP1[30]	LDLRAP1	30	NEG9	30	1.28	1.10	2.39
TOMM40[7.5]	TOMM40	7.5	NEG9	30	1.00	0.30	0.24
SORT1[7.5]	SORT1	7.5	NEG9	30	1.48	0.95	1.52
SORT1[7.5]+TOMM40[7.5]	SORT1	7.5	TOMM40	7.5	0.37	0.33	0.36
MLXIPL[30]+TOMM40[7.5]	MLXIPL	30	TOMM40	7.5	0.70	0.51	0.59
LDLR[30]+LDLRAP1[30]	LDLR	30	LDLRAP1	30	0.25	0.01	0.06
LDLR[30]+MLXIPL[30]	LDLR	30	MLXIPL	30	0.67	0.01	0.03

blot#4	siRNA quantities				LDLR/a-tubulin		
Treatment	**siRNA1**	**pmol1**	**siRNA2**	**pmol2**	**Exp.1**	**Exp.2**	**Exp.3**
NEG9[30]	NEG9	30	NEG9	30	2.26	1.41	0.85
CXCL12[30]	CXCL12	30	NEG9	30	0.86	0.90	0.75
PAFAH1B1[7.5]	PAFAH1B1	7.5	NEG9	30	0.42	0.54	0.51
SIK3[30]	SIK3	30	NEG9	30	1.14	0.75	1.14
MYBPHL[30]	MYBPHL	30	NEG9	30	1.36	1.37	1.82
MYBPHL[30]+SIK3[30]	MYBPHL	30	SIK3	30	0.95	0.82	1.51
PAFAH1B1[7.5]+SIK3[30]	PAFAH1B1	7.5	SIK3	30	0.25	0.30	0.55
CXCL12[30]+PAFAH1B1[7.5]	CXCL12	30	PAFAH1B1	7.5	0.14	0.44	0.30

blot#5	siRNA quantities				LDLR/a-tubulin		
Treatment	**siRNA1**	**pmol1**	**siRNA2**	**pmol2**	**Exp.1**	**Exp.2**	**Exp.3**
NEG9[30]	NEG9	30	NEG9	30	1.60	0.89	1.52
NCAN[30]	NCAN	30	NEG9	30	1.17	0.67	2.83
TOMM40[7.5]	TOMM40	7.5	NEG9	30	0.21	0.38	0.28
LDLR[7.5]	LDLR	7.5	NEG9	30	0.03	0.21	0.10
SEZ6L[15]	SEZ6L	15	NEG9	30	0.53	1.11	1.16
NCAN[30]+SEZ6L[15]	NCAN	30	SEZ6L	15	0.38	1.33	0.70
NCAN[30]+LDLR[7.5]	NCAN	30	LDLR	7.5	0.01	0.20	0.02
NCAN[30]+TOMM40[7.5]	NCAN	30	TOMM40	7.5	0.11	1.03	0.13

blot#6	siRNA quantities				LDLR/a-tubulin		
Treatment	**siRNA1**	**pmol1**	**siRNA2**	**pmol2**	**Exp.1**	**Exp.2**	**Exp.3**
NEG9[30]	NEG9	30	NEG9	30	0.62	1.90	1.18
CELSR2[15]	CELSR2	15	NEG9	30	1.03	1.38	0.71
LPL[15]	LPL	15	NEG9	30	0.44	0.29	0.48
APOB[30]	APOB	30	NEG9	30	1.00	0.68	0.57
HMGCR[30]	HMGCR	30	NEG9	30	1.15	0.96	1.59
TMEM57[15]	TMEM57	15	NEG9	30	1.34	1.43	0.92
PCSK9[15]	PCSK9	15	NEG9	30	1.14	0.47	0.62
PCSK9[15]+TMEM57[15]	PCSK9	15	TMEM57	15	0.91	0.66	1.08
APOB[30]+HMGCR[30]	APOB	30	HMGCR	30	0.76	0.58	1.29
CELSR2[15]+LPL[15]	CELSR2	15	LPL	15	0.42	0.13	0.61

Table S13. Regulation of LDLR transcription and translation upon single and double knockdown of the genes that were identified by the screen as interacting.
Significant effect calculated with Student's t-test in Microsoft Excel (TTEST (array1, array2, 2, 3)) is shown in bold. *<0.05, **<0.01, ***<0.001

Treatment	mRNA						protein				
	log2 (fold change)						LDLR/a-tubulin				
	Exp. 1	Exp. 2	Exp. 3	Exp. 4	median	mRNA	Exp.1	Exp.2	Exp.3	mean	protein
HAVCR1[15]+NEG9	0.23	0.81	1.22	0.66	0.74	*	1.40	0.50	2.90	1.60	
LDLR[7.5]+NEG9	-1.29	-1.50	-1.04	-1.25	-1.27	**	0.10	0.10	0.10	0.10	*
LDLRAP1[30]+NEG9	-0.81	-0.03	0.75	0.07	0.02		0.90	0.30	1.30	0.83	
MLXIPL[7.5]+NEG9	0.75	0.39	0.97	-0.38	0.57		0.80	0.40	2.40	1.20	
BCAM[30]+NEG9	-1.02	-0.69	0.75	-0.03	-0.36		0.50	0.50	0.90	0.63	*
HAVCR1[15]+LDLR[7.5]	-0.57	-1.62	-0.97	-1.76	-1.29	*	0.20	0.00	0.00	0.07	
HAVCR1[15]+LDLRAP1[30]	1.19	1.29	1.43		1.29	*	0.60	0.20	1.10	0.63	
HAVCR1[15]+BCAM[30]	-0.37	-0.60	-0.71		-0.60		0.50	0.50	1.30	0.77	
HAVCR1[15]+MLXIPL[7.5]	0.19	1.87	1.81	-0.42	1.00		0.50	0.30	1.50	0.77	
BCAM[30]+LDLRAP1[30]	-0.17	0.29	0.79	-0.26	0.06		0.70	0.60	0.60	0.63	*
SEZ6L[30]+NEG9	-0.46	0.31	-0.32	-0.46	-0.39		0.60	0.30	0.30	0.40	**
SORT1[30]+NEG9	-0.01	-0.16	-0.38	-0.12	-0.14		0.80	0.10	0.10	0.33	*
LDLR[30]+NEG9	-2.19	-0.40	-2.16	-2.06	-2.11	**	0.10	0.00	0.10	0.07	*
HAVCR1[30]+NEG9	1.05	1.60	1.02	-0.28	1.03			0.80	0.10	0.45	
MLXIPL[30]+NEG9	-0.30	0.38	0.24	-1.86	-0.03		1.30	0.80	0.80	0.97	
LDLR[30]+MLXIPL[30]	-2.40	-2.49	-2.70	-3.13	-2.60	***	0.60	0.00	0.00	0.20	*
LDLR[30]+LDLRAP1[30]	-1.80	-2.15	-1.93	-2.74	-2.04	**	0.20	0.00	0.00	0.07	*
HAVCR1[30]+SEZ6L[30]	0.47	0.13	0.12	0.21	0.17		0.40	0.80	0.50	0.57	*
HAVCR1[30]+SORT1[30]	0.83	0.91	0.22	-0.13	0.52		0.50	0.50	0.20	0.40	**
LDLRAP1[30]+SORT1[30]	-1.30		-0.06	0.47	-0.06		0.80	0.20	0.10	0.37	*
TOMM40[7.5]+NEG9	-0.28	-1.19	-1.02	-0.04	-0.65		0.80	0.10	0.10	0.33	*
MLXIPL[30]+TOMM40[7.5]	0.58	-1.94	-1.02	-0.61	-0.81		0.60	0.20	0.30	0.37	*
CXCL12[30]+NEG9	0.06		1.19	0.21	0.21		0.40	0.60	0.90	0.63	
PAFAH1B1[7.5]+NEG9	-0.91	-1.24	0.47	0.06	-0.42		0.20	0.40	0.60	0.40	
SORT1[7.5]+NEG9	-0.07	-1.36	-0.36	-0.89	-0.63		1.20	0.40	0.70	0.77	
NCAN[30]+NEG9	0.21	0.74	-0.22	-0.38	-0.01		0.70	0.80	1.90	1.13	
CXCL12[30]+PAFAH1B1[7.5]	-0.64	-0.20	0.44	-1.28	-0.42		0.10	0.30	0.40	0.27	*
SORT1[7.5]+TOMM40[7.5]	-0.59	-1.68	-1.05	-1.37	-1.21		0.30	0.20	0.20	0.23	*
NCAN[30]+TOMM40[7.5]	-1.16	-0.92	-1.38	-1.89	-1.27	*	0.10	1.20	0.10	0.47	*
SEZ6L[15]+NEG9	-0.55	0.10	0.49	-0.91	-0.23		0.30	1.30	0.80	0.80	
NCAN[30]+SEZ6L[15]	-1.19	-0.65	0.16	0.08	-0.29		0.20	1.50	0.50	0.73	
LDLR[7.5]+NCAN[30]	-1.98	-2.18	-1.78	-1.31	-1.88	**	0.00	0.20	0.00	0.07	*
SIK3[30]+NEG9	0.23	-0.20	0.59	-0.08	0.08		0.50	0.50	1.30	0.77	
PAFAH1B1[7.5]+SIK3[30]	-1.15	0.38	0.70	0.97	0.54		0.10	0.20	0.60	0.30	
MYBPHL[30]+NEG9	-0.08	0.03	1.33	0.38	0.21		0.60	1.00	2.10	1.23	
MYBPHL[30]+SIK3[30]	0.28	0.90	0.99	0.07	0.59		0.40	0.60	1.80	0.93	
PCSK9[15]+NEG9	-2.14	0.81	0.94	-1.12	-0.15		1.80	0.20	0.50	0.83	
TMEM57[15]+NEG9	-1.06	0.45	1.27	0.34	0.39		2.20	0.80	0.80	1.27	
PCSK9[15]+TMEM57[15]	0.47	-0.19	1.23	0.00	0.24		1.50	0.30	0.90	0.90	
CELSR2[15]+NEG9	-0.11	-0.43	0.91	0.07	-0.02		1.70	0.70	0.60	1.00	
LPL[15]+NEG9	-0.47		0.48	-0.44	-0.44		0.70	0.20	0.40	0.43	
CELSR2[15]+LPL[15]	0.29		0.46	-0.33	0.29		0.70	0.10	0.50	0.43	
APOB[30]+NEG9	0.14	-2.06	1.27	0.37	0.25		1.60	0.40	0.50	0.83	
HMGCR[30]+NEG9	-0.02	-0.76	1.40	0.62	0.30		1.80	0.50	1.30	1.20	
APOB[30]+HMGCR[30]	-1.60	0.37	1.47	0.78	0.57		1.20	0.30	1.10	0.87	

Table S14. Results of qRT-PCR experiments for SREBF1 and SREBF2 mRNA expression used for interaction calling.
Shown are the ΔCt values of SREBF1 and SREBF2 expression, where ΔCt=Ct(SREBF)-Ct(GAPDH), for four biological replicas. The four ΔCt values were used to fit a linear regression model, where the double knockdown was compared with the additive effect of the two single knockdowns, considering NEG9 as control.

					mRNA expression ΔCt					
					SREBF1			SREBF2		
Treatment	siRNA1	pmol	siRNA2	pmol	Exp. 1	Exp. 2	Exp. 3	Exp. 1	Exp. 2	Exp. 3
NEG9[30]	NEG9	30	NEG9	30	7.81	6.79	6.81	5.82	5.72	5.73
HAVCR1[15]+NEG9	HAVCR1	15	NEG9	30	7.83	6.61	6.76	7.04	6.56	5.78
LDLR[7.5]+NEG9	LDLR	7.5	NEG9	30	7.81	6.69	7.43	6.25	5.61	6.52
LDLRAP1[30]+NEG9	LDLRAP1	30	NEG9	30	7.39	6.91	6.53	6.08	5.92	5.40
MLXIPL[7.5]+NEG9	MLXIPL	7.5	NEG9	30	7.25	7.16	7.57	5.62	5.53	5.82
BCAM[30]+NEG9	BCAM	30	NEG9	30	8.15	9.11	7.85	6.40	6.94	6.30
HAVCR1[15]+LDLR[7.5]	HAVCR1	15	LDLR	7.5	8.05	6.12	7.05	7.20	5.50	6.08
HAVCR1[15]+LDLRAP1[30]	HAVCR1	15	LDLRAP1	30	8.05	6.26	7.06	6.75	6.14	6.26
HAVCR1[15]+BCAM[30]	HAVCR1	15	BCAM	30	8.07	7.62	8.12	6.56	7.14	7.14
HAVCR1[15]+MLXIPL[7.5]	HAVCR1	15	MLXIPL	7.5	6.19	7.52	6.52	6.49	6.40	6.47
BCAM[30]+LDLRAP1[30]	BCAM	30	LDLRAP1	30	7.33	7.17	7.77	6.06	6.59	7.00
SEZ6L[30]+NEG9	SEZ6L	30	NEG9	30	8.46	6.54	0.74	7.07	5.75	0.86
SORT1[30]+NEG9	SORT1	30	NEG9	30	4.70	7.04	1.11	3.23	5.57	0.98
LDLR[30]+NEG9	LDLR	30	NEG9	30	7.37	6.95		4.91	5.79	
HAVCR1[30]+NEG9	HAVCR1	30	NEG9	30	6.41	7.27	6.49	6.60	7.66	5.75

MLXIPL[30]+NEG9	MLXIPL	30	NEG9	30	7.28	1.16	9.31	4.93	5.47	7.66
LDLR[30]+MLXIPL[30]	LDLR	30	MLXIPL	30	8.60	7.10	6.91	4.72	5.57	6.00
LDLR[30]+LDLRAP1[30]	LDLR	30	LDLRAP1	30	6.66	8.03	7.36	5.59	6.12	5.60
HAVCR1[30]+SEZ6L[30]	HAVCR1	30	SEZ6L	30	6.83	7.13	5.41	5.29	7.15	5.29
HAVCR1[30]+SORT1[30]	HAVCR1	30	SORT1	30	5.96	5.88	6.73	4.91	5.63	4.80
LDLRAP1[30]+SORT1[30]	LDLRAP1	30	SORT1	30	8.30		7.32	6.39	5.71	6.10
TOMM40[7.5]+NEG9	TOMM40	7.5	NEG9	30	8.68	7.43	7.40	7.45	7.13	6.52
MLXIPL[30]+TOMM40[7.5]	MLXIPL	30	TOMM40	7.5	7.70	9.74	7.53	6.90	8.02	6.27
CXCL12[30]+NEG9	CXCL12	30	NEG9	30	6.83	4.16	7.78	6.24	5.77	5.94
PAFAH1B1[7.5]+NEG9	PAFAH1B1	7.5	NEG9	30	6.19	6.50	6.83	5.71	6.26	5.84
SORT1[7.5]+NEG9	SORT1	7.5	NEG9	30	6.49	6.85	7.24	6.28	14.72	5.78
NCAN[30]+NEG9	NCAN	30	NEG9	30	7.81	7.59	8.35	7.50	6.35	6.52
CXCL12[30]+PAFAH1B1[7.5]	CXCL12	30	PAFAH1B1	7.5	6.48	6.91	8.05	5.06	6.07	6.13
SORT1[7.5]+TOMM40[7.5]	SORT1	7.5	TOMM40	7.5	6.88	7.94	7.92	6.36	6.69	6.52
NCAN[30]+TOMM40[7.5]	NCAN	30	TOMM40	7.5	7.60	7.68	8.78	5.78	6.95	6.70
SEZ6L[15]+NEG9	SEZ6L	15	NEG9	30	7.71	7.83	7.37	5.74	7.39	6.17
NCAN[30]+SEZ6L[15]	NCAN	30	SEZ6L	15	7.70	7.94	8.25	6.92	6.37	6.08
LDLR[7.5]+NCAN[30]	LDLR	7.5	NCAN	30	7.70	7.60	7.77	5.72	5.99	6.26
SIK3[30]+NEG9	SIK3	30	NEG9	30	8.11	7.65	7.97	5.88	6.56	7.61
PAFAH1B1[7.5]+SIK3[30]	PAFAH1B1	7.5	SIK3	30	9.35	8.05	8.37	8.80	7.56	7.38
MYBPHL[30]+NEG9	MYBPHL	30	NEG9	30	6.47	6.61	7.19	5.18	5.39	6.15
MYBPHL[30]+SIK3[30]	MYBPHL	30	SIK3	30	7.04	7.36	7.26	6.84	6.89	6.63
PCSK9[15]+NEG9	PCSK9	15	NEG9	30		7.78	7.21	7.75	6.01	6.29
TMEM57[15]+NEG9	TMEM57	15	NEG9	30	7.31	7.80	7.89	6.16	6.80	6.87
PCSK9[15]+TMEM57[15]	PCSK9	15	TMEM57	15	6.74	6.49	6.78	5.55	5.85	6.16
CELSR2[15]+NEG9	CELSR2	30	NEG9	30	7.61	6.88	7.51	5.83	5.83	5.98
LPL[15]+NEG9	LPL	30	NEG9	30	9.18	5.39	8.38	6.75		6.33
CELSR2[15]+LPL[15]	CELSR2	30	LPL	30	7.36	n.d.	7.67	6.29	8.95	6.78
APOB[30]+NEG9	APOB	30	NEG9	30	5.98	8.31	6.23	4.62	8.33	5.85
HMGCR[30]+NEG9[30]	HMGCR	30	NEG9	30	7.81	7.51	7.40	4.73	6.43	6.72
APOB[30]+HMGCR[30]	APOB	30	HMGCR	30	7.13	8.66	7.36	4.31	6.56	5.90

Table S15. Regulation of SREBF1 and SREBF2 transcription upon single and double knockdown of the genes that were identified by the screen as interacting.

Significant effect calculated with Student's t-test in Microsoft Excel (TTEST (array1, array2, 2, 3)) is shown in bold. *<0.1, **<0.01, ***<0.001.

	SREBF1					SREBF2				
	log2 (fold change)					log2 (fold change)				
Treatment	**Exp.1**	**Exp.2**	**Exp.3**	**mean**	**mRNA**	**Exp.1**	**Exp.2**	**Exp.3**	**mean**	**mRNA**
HAVCR1[15]+NEG9	-0.02	0.18	0.05	0.07		-1.21	-0.84	-0.05	-0.70	
LDLR[7.5]+NEG9	0.00	0.11	-0.63	-0.17		-0.42	0.12	-0.79	-0.36	
LDLRAP1[30]+NEG9	0.42	-0.12	0.28	0.19		-0.25	-0.20	0.33	-0.04	
MLXIPL[7.5]+NEG9	0.58	-0.36	-0.77	-0.18			0.19	-0.09	0.05	
BCAM[30]+NEG9	-0.34	-2.31	-1.05	-1.23	*	-0.16	-1.22	-0.57	-0.65	*
HAVCR1[15]+LDLR[7.5]	-0.66	0.68	-0.25	-0.08		-1.12	0.22	-0.35	-0.42	
HAVCR1[15]+LDLRAP1[30]	-0.80	0.53	-0.26	-0.18		-1.13	-0.42	-0.53	-0.69	*
HAVCR1[15]+BCAM[30]	0.08	-0.83	-1.32	-0.69		-0.16	-1.42	-1.41	-0.99	*
HAVCR1[15]+MLXIPL[7.5]		-0.72	0.28	-0.22		0.71	-0.67	-0.74	-0.24	***
BCAM[30]+LDLRAP1[30]	0.72	-0.38	-0.96	-0.20		0.69	-0.86	-1.27	-0.48	*
SEZ6L[30]+NEG9	-0.39	0.25	-0.43	-0.19		-0.51	-0.03	-0.22	-0.26	
SORT1[30]+NEG9		-0.24	0.15	-0.04			0.15	-0.03	0.06	
LDLR[30]+NEG9	-0.04	-0.16		-0.10			-0.07		-0.07	
HAVCR1[30]+NEG9		-0.48	0.66	0.09		0.47	-1.94	0.11	-0.45	
MLXIPL[30]+NEG9	-2.59		-2.17	-2.38		-1.70		-1.80	-1.75	
LDLR[30]+MLXIPL[30]	-1.22	-0.30	0.27	-0.42		0.19	0.15	0.29	0.21	
LDLR[30]+LDLRAP1[30]	-0.24	-1.23	-0.17	-0.55		1.00	-0.40	0.69	0.43	
HAVCR1[30]+SEZ6L[30]	0.45	-0.33		0.06		-0.35	-1.43	0.57	-0.40	
HAVCR1[30]+SORT1[30]		0.92	0.42	0.67		-0.19	0.09	1.06	0.32	
LDLRAP1[30]+SORT1[30]	-1.65		-0.14	-0.89		-0.79	0.01	0.19	-0.20	
TOMM40[7.5]+NEG9	-1.85	-0.64	-0.25	-0.91		-2.16	-1.41	-0.66	-1.41	*
MLXIPL[30]+TOMM40[7.5]	-1.74	-2.95	-0.38	-1.69		-1.99	-2.30	-0.41	-1.57	
CXCL12[30]+NEG9			-0.98	-0.98		0.15	-0.05	-0.21	-0.04	
PAFAH1B1[7.5]+NEG9		0.30	-0.03	0.14			-0.54	-0.11	-0.33	
SORT1[7.5]+NEG9		-0.05	-0.43	-0.24		0.63		-0.05	0.29	
NCAN[30]+NEG9	-0.98	-0.79	-1.54	-1.10		-1.26	-0.63	-0.79	-0.89	
CXCL12[30]+PAFAH1B1[7.5]	-0.28	-0.11	-1.24	-0.55		0.66	-0.35	-0.40	-0.03	
SORT1[7.5]+TOMM40[7.5]	-0.39	-1.15	-1.12	-0.89		-0.09	-0.97	-0.79	-0.61	**
NCAN[30]+TOMM40[7.5]	0.22	-0.88	-1.97	-0.88			-1.23	-0.97	-1.10	*
SEZ6L[15]+NEG9	-1.24	-1.03	-0.57	-0.95		-0.68	-1.67	-0.44	-0.93	
NCAN[30]+SEZ6L[15]	-0.81	-1.15	-1.45	-1.14		-0.56	-0.65	-0.35	-0.52	
LDLR[7.5]+NCAN[30]	-0.10	-0.81	-0.97	-0.63		0.05	-0.27	-0.53	-0.25	
SIK3[30]+NEG9	-0.40	-0.85	-1.17	-0.81		-0.14	-0.84	-1.88	-0.95	
PAFAH1B1[7.5]+SIK3[30]	-1.66	-1.25	-1.56	-1.49	*	-1.88	-1.84	-1.65	-1.79	*
MYBPHL[30]+NEG9		0.19	-0.39	-0.10		0.54	0.33	-0.42	0.15	
MYBPHL[30]+SIK3[30]		-0.56	-0.46	-0.51	*	-0.96	-1.17	-0.90	-1.01	**
PCSK9[15]+NEG9		-0.99	-0.40	-0.70			-0.29	-0.56	-0.43	
TMEM57[15]+NEG9	-0.83	-1.01	-1.08	-0.97		-0.97	-1.08	-1.14	-1.06	*
PCSK9[15]+TMEM57[15]	0.31	0.31	0.03	0.21			-0.13	-0.43	-0.28	

CELSR2[15]+NEG9		-0.09	-0.70	-0.39		-0.11	-0.25	-0.18
LPL[15]+NEG9	-1.87		-1.58	-1.72	-0.60		-0.60	-0.60
CELSR2[15]+LPL[15]	-0.31		-0.86	-0.59	0.44		-1.05	-0.30
APOB[30]+NEG9	1.07		0.57	0.82	0.61		-0.12	0.25
HMGCR[30]+NEG9[30]	-0.77	-1.12	-0.60	-0.83		-1.29	-0.99	-1.14
APOB[30]+HMGCR[30]	-0.09	-2.27	-0.55	-0.97	0.92	-1.42	-0.17	-0.22

Table S16. Effect of the gene interactions on LDLR mRNA and protein levels.

Interaction effects on LDLR mRNA and protein of the 21 validated gene interactions, calculated by fitting a linear model on the Dct values of 4 biological replicates and the LDLR/a-tubulin protein values of 3 biological replicates, whereby the effect of single knockdowns was compared to that of doubles. The standard error, as well as the p-value of the interaction are shown. The significant interactions (pVal<10^{-1}) are highlighted in bold.

		LDLR mRNA			LDLR protein		
Gene1	**Gene2**	**Interaction Effect**	**Std.Error**	**pvalue**	**Interaction Effect**	**Std.Error**	**pvalue**
HAVCR	**LDLR**	**0.91**	**0.45**	**0.04**	**-0.44**	**0.24**	**0.07**
HAVCR1	LDLRAP	-0.43	0.43	0.32	-0.70	0.54	0.19
HAVCR1	BCAM	-0.43	0.43	0.32	-0.29	0.66	0.66
HAVCR1	MLXIPL	0.10	0.52	0.86	**0.51**	**0.26**	**0.05**
BCAM	LDLRAP	-0.27	0.58	0.64	0.38	0.41	0.36
LDLR	**MLXIPL**	**1.00**	**0.52**	**0.05**	0.42	0.32	0.19
LDLR	LDLRAP	0.65	0.57	0.26	0.30	0.38	0.43
HAVCR	**SEZ6L**	**0.78**	**0.47**	**0.10**	0.32	0.22	0.14
HAVCR1	SORT1	0.53	0.48	0.27	**0.67**	**0.22**	**0.00**
LDLRAP	SORT1	0.15	0.65	0.82	**0.59**	**0.23**	**0.01**
TOMM4	MLXIPL	0.13	0.93	0.89	0.50	0.48	0.30
CXCL12	PAFAH1	0.77	0.61	0.20	**0.38**	**0.22**	**0.09**
SORT1	TOMM4	0.23	0.94	0.81	0.27	0.31	0.38
NCAN	**TOMM**	**1.16**	**0.51**	**0.02**	-0.03	0.56	0.96
SEZ6L	NCAN	0.55	0.43	0.20	0.25	0.58	0.66
LDLR	**NCAN**	**0.86**	**0.49**	**0.08**	0.27	0.24	0.26
MYBPH	SIK3	0.08	0.53	0.88	-0.05	0.62	0.94
PAFAH1	SIK3	-0.31	0.70	0.66	0.13	0.46	0.78
PCSK9	TMEM5	-0.31	0.70	0.66	0.09	0.48	0.85
CELSR2	LPL	-0.13	0.66	0.85	0.08	0.56	0.89
APOB	HMGCR	0.45	1.62	0.78	-0.88	0.52	0.09

Table S17. Effect of the gene interactions on SREBF1 and SREBF2 mRNA levels.

Shown are the interaction effects of the 21 validated gene interactions on SREBF1 and SREBF2 mRNA, as calculated by fitting a linear model on the DCt values of 4 biological replicates, to compare the effect of the single knockdowns to that of the doubles. The standard error of the estimated interaction effect, as well as the p-value of the interaction are shown. The significant ones (pVal<10^{-1}) are highlighted in bold.

		SREBF1 mRNA			SREBF2 mRNA		
Gene1	**Gene2**	**Interaction Effect**	**Std.Error**	**pvalue**	**Interaction Effect**	**Std.Error**	**pvalue**
HAVCR1	LDLR	-0.18	0.92	0.85	**-1.08**	**0.28**	**0.00**
HAVCR1	LDLRAP1	0.22	0.85	0.80	-0.85	0.85	0.32
HAVCR1	BCAM	-0.27	0.88	0.76	-0.35	0.42	0.40
HAVCR1	MLXIPL	-0.20	0.74	0.79	-0.15	0.22	0.51
BCAM	LDLRAP1	-0.73	0.65	0.26	-0.03	0.22	0.90
LDLR	MLXIPL	-0.81	1.29	0.53	-0.01	0.39	0.99
LDLR	LDLRAP1	0.38	0.76	0.61	**-0.76**	**0.45**	**0.09**
HAVCR1	SEZ6L	0.69	1.06	0.51	**-0.32**	**0.16**	**0.05**
HAVCR1	**SORT1**	**2.48**	**1.49**	**0.09**	-1.19	0.48	0.01
LDLRAP1	**SORT1**	**3.93**	**1.32**	**0.00**	1.00	1.25	0.42
TOMM40	MLXIPL	-0.71	1.35	0.60	**2.75**	**0.65**	**0.00**
CXCL12	PAFAH1B1	-0.46	0.91	0.61	-0.57	1.28	0.66
SORT1	TOMM40	-0.07	0.90	0.94	-0.39	0.37	0.29
NCAN	TOMM40	-0.57	0.76	0.46	**-0.88**	**0.38**	**0.02**
SEZ6L	NCAN	-0.53	0.46	0.25	**-1.11**	**0.38**	**0.00**
LDLR	NCAN	-0.42	0.57	0.46	-0.63	0.51	0.22
MYBPHL	SIK3	-0.45	0.50	0.37	**-1.07**	**0.36**	**0.00**
PAFAH1B1	SIK3	0.83	0.52	0.11	**0.60**	**0.29**	**0.04**
PCSK9	**TMEM57**	**-1.47**	**0.45**	**0.00**	0.84	0.63	0.19
CELSR2	**LPL**	**-1.35**	**0.77**	**0.08**	-0.15	0.50	0.76
APOB	HMGCR	0.71	0.44	0.10	-0.77	0.56	0.17

Table S18. Correlation of LDLR mRNA and protein, as well as SREBF1 and SREBF2 mRNA with functional effect on LDL uptake.
Shown is the effect on LDL uptake, LDLR mRNA and protein, as well as SREBF1 and SREBF2 mRNA, upon double knockdowns of the genes which were identified by the screen as interacting. Interaction effects are highlighted with pink (increased) or green (decreased).

		LDL uptake	LDLR mRNA	LDLR protein	SREBF1 mRNA	SREBF2 mRNA	correlation	anti-correlation
	Treatment	**Robust Z score**	**log2 (fold change)**	**log2(LDLR/a-tubulin)**	**log2 (fold change)**	**log2 (fold change)**		
1	LDLR+ NCAN	**-1.23**	**-1.88**	-3.91	-0.63	**-0.25**	LDL-LDLR mRNA-SREBF2 mRNA	
2	CXCL12+ PAFAH1B1	**1.32**	-0.42	**-1.91**	-0.55	-0.03		LDL-LDLR protein
3	NCAN + SEZ6L	**-0.08**	-0.29	-0.45	-1.14	-0.52		
4	NCAN + TOMM40	**-1.49**	**-1.27**	-1.10	-0.88	**-1.10**	LDL-LDLR mRNA-SREBF2 mRNA	
5	HAVCR1 + MLXIPL	**-0.12**	1.00	-0.38	-0.22	-0.24		
6	LDLRAP1 + SORT1	**-1.09**	-0.06	-1.45	**-0.89**	**-0.20**	LDL-SREBF1 mRNA-SREBF2	
7	LDLR + LDLRAP1	**-2.49**	-2.04	**-3.91**	-0.55	**0.43**	LDL-LDLR protein	LDL-SREBF2
8	CELSR2 + LPL	**-0.11**	0.29	-1.21	**-0.59**	-1.28	LDL-SREBF1 mRNA	
9	MLXIPL + TOMM40	**-1.94**	-0.81	**-1.45**	-1.69	-1.57	LDL-LDLR protein	
10	LDLR + MLXIPL	**-2.26**	**-2.60**	-2.32	-0.42	**0.21**	LDL-LDLR mRNA	LDL-SREBF2
11	HAVCR1+ SEZ6L	**-0.67**	**0.17**	-0.82	0.06	-0.40		
12	HAVCR1 + LDLRAP1	**0.39**	1.29	-0.66	-0.18	-0.69		
13	BCAM + LDLRAP1	**-0.05**	0.06	**-0.66**	-0.20	-0.48	LDL-LDLR protein	
14	HAVCR1 + SORT1	**-0.63**	0.52	-1.32	**0.67**	0.32		LDL-SREBF1
15	BCAM + HAVCR1	**-0.51**	-0.60	-0.38	-0.69	-0.99		
16	MYBPHL + SIK3	**2.30**	0.59	-0.10	-0.51	**-1.01**		LDL-SREBF2
17	APOB + HMGCR	**3.33**	0.57	-0.21	-0.97	0.25		
18	SORT1 + TOMM40	**-2.48**	-1.21	-2.10	-0.89	**-0.61**	LDL-SREBF2 mRNA	
19	PAFAH1B1+ SIK3	**3.30**	0.54	-1.74	-1.49	-1.79		
20	PCSK9 + TMEM57	**3.21**	0.24	-0.15	**0.21**	**-0.28**	LDL-SREBF1 mRNA	LDL-SREBF2
21	HAVCR1+LDLR	**-2.24**	**-1.29**	**-3.91**	-0.08	-0.42	LDL-LDLR mRNA-LDLR protein	

Table S19. Results of the co-RNAi screen.

Treatment	siRNA1	siRNA2	robustZscore	mad
APOB__APOB	APOB	APOB	-0.84	0.49
APOB__APOE	APOB	APOE	0.29	1.00
APOB__BAZ1B	APOB	BAZ1B	0.41	0.78
APOB__BCAM	APOB	BCAM	-0.07	0.82
APOB__BCL7B	APOB	BCL7B	0.10	0.82
APOB__CBLC	APOB	CBLC	-0.09	1.24
APOB__CELSR2	APOB	CELSR2	0.83	0.79
APOB__CXCL12	APOB	CXCL12	-0.05	2.02
APOB__FAM174A	APOB	FAM174A	-1.40	0.97
APOB__HAVCR1	APOB	HAVCR1	-0.04	1.08
APOB__HMGCR	APOB	HMGCR	2.80	1.53
APOB__LDLR	APOB	LDLR	-1.68	0.96
APOB__LDLRAP1	APOB	LDLRAP1	0.30	0.86
APOB__LPL	APOB	LPL	-0.86	0.69
APOB__MLXIPL	APOB	MLXIPL	-0.46	0.68
APOB__MYBPHL	APOB	MYBPHL	0.58	0.62
APOB__MYLIP	APOB	MYLIP	0.38	1.23
APOB__NCAN	APOB	NCAN	-0.30	0.97
APOB__NEG9	APOB	NEG9	-0.10	1.10
APOB__PAFAH1B1	APOB	PAFAH1B1	-0.13	1.54
APOB__PAFAH1B2	APOB	PAFAH1B2	-0.27	0.94
APOB__PCSK9	APOB	PCSK9	-0.74	1.49
APOB__PVRL2	APOB	PVRL2	0.03	0.78
APOB__SEZ6L	APOB	SEZ6L	-0.61	0.50
APOB__SIK3	APOB	SIK3	2.16	1.31
APOB__SORT1	APOB	SORT1	-0.29	0.44
APOB__TM6SF2	APOB	TM6SF2	-0.32	1.33
APOB__TMEM57	APOB	TMEM57	-0.27	0.98
APOB__TOMM40	APOB	TOMM40	-1.26	0.74
APOB__WDR12	APOB	WDR12	0.10	1.78
APOB__ZNF259	APOB	ZNF259	0.01	0.78
APOE__APOE	APOE	APOE	-0.26	0.62
APOE__BAZ1B	APOE	BAZ1B	0.99	1.43
APOE__BCAM	APOE	BCAM	0.28	1.17
APOE__BCL7B	APOE	BCL7B	-0.14	0.78
APOE__CBLC	APOE	CBLC	0.52	0.56
APOE__CELSR2	APOE	CELSR2	0.23	1.05
APOE__CXCL12	APOE	CXCL12	1.75	1.77
APOE__FAM174A	APOE	FAM174A	-0.20	0.67
APOE__HAVCR1	APOE	HAVCR1	-0.15	1.48
APOE__HMGCR	APOE	HMGCR	0.96	0.71
APOE__LDLR	APOE	LDLR	-1.07	0.91
APOE__LDLRAP1	APOE	LDLRAP1	0.47	0.95
APOE__LPL	APOE	LPL	-0.10	1.02
APOE__MLXIPL	APOE	MLXIPL	0.35	0.46
APOE__MYBPHL	APOE	MYBPHL	0.55	1.77
APOE__MYLIP	APOE	MYLIP	0.39	0.69
APOE__NCAN	APOE	NCAN	1.57	1.72
APOE__NEG9	APOE	NEG9	0.40	1.29
APOE__PAFAH1B1	APOE	PAFAH1B1	-1.04	1.23
APOE__PAFAH1B2	APOE	PAFAH1B2	0.53	0.95
APOE__PCSK9	APOE	PCSK9	-0.09	1.58
APOE__PVRL2	APOE	PVRL2	1.10	0.89
APOE__SEZ6L	APOE	SEZ6L	-1.00	0.74
APOE__SIK3	APOE	SIK3	0.93	0.87
APOE__SORT1	APOE	SORT1	0.36	0.73
APOE__TM6SF2	APOE	TM6SF2	-0.04	0.72
APOE__TMEM57	APOE	TMEM57	0.23	0.72
APOE__TOMM40	APOE	TOMM40	-0.62	0.11
APOE__WDR12	APOE	WDR12	0.27	0.68
APOE__ZNF259	APOE	ZNF259	0.11	1.53
BAZ1B__BAZ1B	BAZ1B	BAZ1B	-0.98	1.31
BAZ1B__BCAM	BAZ1B	BCAM	-0.63	0.57
BAZ1B__BCL7B	BAZ1B	BCL7B	0.47	1.40
BAZ1B__CBLC	BAZ1B	CBLC	0.31	0.69
BAZ1B__CELSR2	BAZ1B	CELSR2	0.39	0.44
BAZ1B__CXCL12	BAZ1B	CXCL12	0.41	1.58
BAZ1B__FAM174A	BAZ1B	FAM174A	-0.59	1.22
BAZ1B__HAVCR1	BAZ1B	HAVCR1	0.30	0.63
BAZ1B__HMGCR	BAZ1B	HMGCR	1.00	1.39
BAZ1B__LDLR	BAZ1B	LDLR	-1.65	1.02
BAZ1B__LDLRAP1	BAZ1B	LDLRAP1	0.52	0.88
BAZ1B__LPL	BAZ1B	LPL	-0.69	0.97
BAZ1B__MLXIPL	BAZ1B	MLXIPL	0.53	0.82
BAZ1B__MYBPHL	BAZ1B	MYBPHL	0.22	1.21
BAZ1B__MYLIP	BAZ1B	MYLIP	0.77	1.01
BAZ1B__NCAN	BAZ1B	NCAN	1.26	2.15
BAZ1B__NEG9	BAZ1B	NEG9	0.42	1.65
BAZ1B__PAFAH1B1	BAZ1B	PAFAH1B1	-1.08	1.60
BAZ1B__PAFAH1B2	BAZ1B	PAFAH1B2	-0.31	0.64
BAZ1B__PCSK9	BAZ1B	PCSK9	0.25	0.92
BAZ1B__PVRL2	BAZ1B	PVRL2	-0.30	1.94
BAZ1B__SEZ6L	BAZ1B	SEZ6L	-0.30	0.36
BAZ1B__SIK3	BAZ1B	SIK3	0.12	0.58
BAZ1B__SORT1	BAZ1B	SORT1	-0.57	0.36
BAZ1B__TM6SF2	BAZ1B	TM6SF2	-0.66	0.43
BAZ1B__TMEM57	BAZ1B	TMEM57	1.35	1.51
BAZ1B__TOMM40	BAZ1B	TOMM40	-1.32	0.22
BAZ1B__WDR12	BAZ1B	WDR12	0.32	0.97
BAZ1B__ZNF259	BAZ1B	ZNF259	-0.13	0.94
BCAM__BCAM	BCAM	BCAM	-0.37	0.45
BCAM__BCL7B	BCAM	BCL7B	-0.43	0.59
BCAM__CBLC	BCAM	CBLC	-0.26	0.87
BCAM__CELSR2	BCAM	CELSR2	0.69	1.03
BCAM__CXCL12	BCAM	CXCL12	0.17	0.67
BCAM__FAM174A	BCAM	FAM174A	-0.49	1.25
BCAM__HAVCR1	BCAM	HAVCR1	-0.51	1.03

BCAM__HMGCR	BCAM	HMGCR	1.27	1.19
BCAM__LDLR	BCAM	LDLR	-2.05	0.84
BCAM__LDLRAP1	BCAM	LDLRAP1	-0.40	0.88
BCAM__LPL	BCAM	LPL	-0.74	0.55
BCAM__MLXIPL	BCAM	MLXIPL	0.44	0.66
BCAM__MYBPHL	BCAM	MYBPHL	-0.03	0.73
BCAM__MYLIP	BCAM	MYLIP	0.63	0.72
BCAM__NCAN	BCAM	NCAN	0.71	0.88
BCAM__NEG9	BCAM	NEG9	0.43	1.51
BCAM__PAFAH1B1	BCAM	PAFAH1B1	-0.34	0.94
BCAM__PAFAH1B2	BCAM	PAFAH1B2	-0.37	0.62
BCAM__PCSK9	BCAM	PCSK9	0.63	0.92
BCAM__PVRL2	BCAM	PVRL2	0.80	1.81
BCAM__SEZ6L	BCAM	SEZ6L	-0.41	1.44
BCAM__SIK3	BCAM	SIK3	-0.25	0.82
BCAM__SORT1	BCAM	SORT1	-0.44	0.84
BCAM__TM6SF2	BCAM	TM6SF2	-0.73	0.32
BCAM__TMEM57	BCAM	TMEM57	0.31	0.39
BCAM__TOMM40	BCAM	TOMM40	-0.78	0.16
BCAM__WDR12	BCAM	WDR12	-0.26	0.62
BCAM__ZNF259	BCAM	ZNF259	-0.07	0.61
BCL7B__BCL7B	BCL7B	BCL7B	2.11	1.02
BCL7B__CBLC	BCL7B	CBLC	-0.64	0.56
BCL7B__CELSR2	BCL7B	CELSR2	1.41	1.59
BCL7B__CXCL12	BCL7B	CXCL12	0.04	0.69
BCL7B__FAM174A	BCL7B	FAM174A	-1.37	0.63
BCL7B__HAVCR1	BCL7B	HAVCR1	0.59	0.94
BCL7B__HMGCR	BCL7B	HMGCR	0.94	0.45
BCL7B__LDLR	BCL7B	LDLR	-1.91	0.79
BCL7B__LDLRAP1	BCL7B	LDLRAP1	0.89	1.41
BCL7B__LPL	BCL7B	LPL	0.23	1.08
BCL7B__MLXIPL	BCL7B	MLXIPL	-0.03	1.35
BCL7B__MYBPHL	BCL7B	MYBPHL	0.46	1.05
BCL7B__MYLIP	BCL7B	MYLIP	0.69	1.33
BCL7B__NCAN	BCL7B	NCAN	0.20	0.85
BCL7B__NEG9	BCL7B	NEG9	0.29	1.23
BCL7B__PAFAH1B1	BCL7B	PAFAH1B1	0.02	0.66
BCL7B__PAFAH1B2	BCL7B	PAFAH1B2	0.57	1.33
BCL7B__PCSK9	BCL7B	PCSK9	-0.85	1.59
BCL7B__PVRL2	BCL7B	PVRL2	0.25	0.94
BCL7B__SEZ6L	BCL7B	SEZ6L	-0.87	0.66
BCL7B__SIK3	BCL7B	SIK3	1.53	1.72
BCL7B__SORT1	BCL7B	SORT1	-0.56	1.57
BCL7B__TM6SF2	BCL7B	TM6SF2	-0.55	0.75
BCL7B__TMEM57	BCL7B	TMEM57	0.26	1.45
BCL7B__TOMM40	BCL7B	TOMM40	-0.12	0.21
BCL7B__WDR12	BCL7B	WDR12	-0.47	0.71
BCL7B__ZNF259	BCL7B	ZNF259	-0.04	1.03
CBLC__CBLC	CBLC	CBLC	-0.13	0.79
CBLC__CELSR2	CBLC	CELSR2	0.67	1.70
CBLC__CXCL12	CBLC	CXCL12	0.36	0.58
CBLC__FAM174A	CBLC	FAM174A	-0.98	0.45
CBLC__HAVCR1	CBLC	HAVCR1	-0.14	0.74
CBLC__HMGCR	CBLC	HMGCR	0.58	0.52
CBLC__LDLR	CBLC	LDLR	-1.60	1.21
CBLC__LDLRAP1	CBLC	LDLRAP1	0.57	1.12
CBLC__LPL	CBLC	LPL	-0.84	0.54
CBLC__MLXIPL	CBLC	MLXIPL	0.41	0.58
CBLC__MYBPHL	CBLC	MYBPHL	0.13	1.29
CBLC__MYLIP	CBLC	MYLIP	0.69	1.73
CBLC__NCAN	CBLC	NCAN	0.78	0.64
CBLC__NEG9	CBLC	NEG9	0.25	1.25
CBLC__PAFAH1B1	CBLC	PAFAH1B1	-0.57	1.34
CBLC__PAFAH1B2	CBLC	PAFAH1B2	0.70	1.05
CBLC__PCSK9	CBLC	PCSK9	-0.19	0.97
CBLC__PVRL2	CBLC	PVRL2	1.10	2.54
CBLC__SEZ6L	CBLC	SEZ6L	-0.97	0.65
CBLC__SIK3	CBLC	SIK3	1.04	0.99
CBLC__SORT1	CBLC	SORT1	-0.59	0.56
CBLC__TM6SF2	CBLC	TM6SF2	-0.07	1.06
CBLC__TMEM57	CBLC	TMEM57	-0.17	0.50
CBLC__TOMM40	CBLC	TOMM40	-0.54	0.73
CBLC__WDR12	CBLC	WDR12	0.63	1.13
CBLC__ZNF259	CBLC	ZNF259	-0.42	0.69
CELSR2__CELSR2	CELSR2	CELSR2	-0.09	1.56
CELSR2__CXCL12	CELSR2	CXCL12	0.86	0.73
CELSR2__FAM174A	CELSR2	FAM174A	-1.11	1.46
CELSR2__HAVCR1	CELSR2	HAVCR1	0.49	1.39
CELSR2__HMGCR	CELSR2	HMGCR	0.63	0.67
CELSR2__LDLR	CELSR2	LDLR	-2.13	1.32
CELSR2__LDLRAP1	CELSR2	LDLRAP1	1.04	1.22
CELSR2__LPL	CELSR2	LPL	-1.40	0.58
CELSR2__MLXIPL	CELSR2	MLXIPL	0.49	0.99
CELSR2__MYBPHL	CELSR2	MYBPHL	1.28	1.89
CELSR2__MYLIP	CELSR2	MYLIP	0.49	0.31
CELSR2__NCAN	CELSR2	NCAN	0.89	1.28
CELSR2__NEG9	CELSR2	NEG9	0.23	1.11
CELSR2__PAFAH1B1	CELSR2	PAFAH1B1	0.29	1.18
CELSR2__PAFAH1B2	CELSR2	PAFAH1B2	0.82	2.30
CELSR2__PCSK9	CELSR2	PCSK9	-0.21	1.16
CELSR2__PVRL2	CELSR2	PVRL2	1.50	1.10
CELSR2__SEZ6L	CELSR2	SEZ6L	-0.97	0.69
CELSR2__SIK3	CELSR2	SIK3	2.30	0.96
CELSR2__SORT1	CELSR2	SORT1	-0.51	0.75
CELSR2__TM6SF2	CELSR2	TM6SF2	-0.63	0.97
CELSR2__TMEM57	CELSR2	TMEM57	0.42	1.16
CELSR2__TOMM40	CELSR2	TOMM40	-0.05	0.74
CELSR2__WDR12	CELSR2	WDR12	0.46	1.28

CELSR2__ZNF259	CELSR2	ZNF259	-0.61	1.39
CXCL12__CXCL12	CXCL12	CXCL12	2.20	1.80
CXCL12__FAM174A	CXCL12	FAM174A	-1.69	0.88
CXCL12__HAVCR1	CXCL12	HAVCR1	-0.41	0.95
CXCL12__HMGCR	CXCL12	HMGCR	1.08	1.04
CXCL12__LDLR	CXCL12	LDLR	-2.15	0.50
CXCL12__LDLRAP1	CXCL12	LDLRAP1	1.37	1.46
CXCL12__LPL	CXCL12	LPL	-1.01	1.16
CXCL12__MLXIPL	CXCL12	MLXIPL	0.71	1.55
CXCL12__MYBPHL	CXCL12	MYBPHL	-0.41	0.64
CXCL12__MYLIP	CXCL12	MYLIP	-0.06	0.74
CXCL12__NCAN	CXCL12	NCAN	1.83	1.14
CXCL12__NEG9	CXCL12	NEG9	0.23	1.39
CXCL12__PAFAH1B1	CXCL12	PAFAH1B1	-2.17	0.71
CXCL12__PAFAH1B2	CXCL12	PAFAH1B2	-0.44	1.01
CXCL12__PCSK9	CXCL12	PCSK9	0.82	1.40
CXCL12__PVRL2	CXCL12	PVRL2	-0.20	1.28
CXCL12__SEZ6L	CXCL12	SEZ6L	-0.78	1.46
CXCL12__SIK3	CXCL12	SIK3	0.01	1.91
CXCL12__SORT1	CXCL12	SORT1	-1.35	1.11
CXCL12__TM6SF2	CXCL12	TM6SF2	-0.97	1.76
CXCL12__TMEM57	CXCL12	TMEM57	0.86	2.16
CXCL12__TOMM40	CXCL12	TOMM40	-1.54	0.80
CXCL12__WDR12	CXCL12	WDR12	-0.04	0.98
CXCL12__ZNF259	CXCL12	ZNF259	-0.69	0.85
FAM174A__FAM174A	FAM174A	FAM174A	-2.32	0.85
FAM174A__HAVCR1	FAM174A	HAVCR1	-1.04	1.26
FAM174A__HMGCR	FAM174A	HMGCR	0.16	0.59
FAM174A__LDLR	FAM174A	LDLR	-2.17	0.41
FAM174A__LDLRAP1	FAM174A	LDLRAP1	-0.77	1.16
FAM174A__LPL	FAM174A	LPL	-1.87	0.74
FAM174A__MLXIPL	FAM174A	MLXIPL	0.58	1.75
FAM174A__MYBPHL	FAM174A	MYBPHL	-0.91	1.51
FAM174A__MYLIP	FAM174A	MYLIP	-0.81	0.95
FAM174A__NCAN	FAM174A	NCAN	0.08	1.43
FAM174A__NEG9	FAM174A	NEG9	-0.78	1.37
FAM174A__PAFAH1B1	FAM174A	PAFAH1B1	-0.60	1.39
FAM174A__PAFAH1B2	FAM174A	PAFAH1B2	-0.96	0.80
FAM174A__PCSK9	FAM174A	PCSK9	-0.53	1.00
FAM174A__PVRL2	FAM174A	PVRL2	-1.72	0.79
FAM174A__SEZ6L	FAM174A	SEZ6L	-0.63	1.55
FAM174A__SIK3	FAM174A	SIK3	0.19	0.99
FAM174A__SORT1	FAM174A	SORT1	-0.22	0.99
FAM174A__TM6SF2	FAM174A	TM6SF2	-1.20	1.08
FAM174A__TMEM57	FAM174A	TMEM57	-0.71	1.97
FAM174A__TOMM40	FAM174A	TOMM40	-1.23	1.41
FAM174A__WDR12	FAM174A	WDR12	-0.83	0.54
FAM174A__ZNF259	FAM174A	ZNF259	-0.96	1.05
HAVCR1__HAVCR1	HAVCR1	HAVCR1	-0.67	1.52
HAVCR1__HMGCR	HAVCR1	HMGCR	0.28	1.38
HAVCR1__LDLR	HAVCR1	LDLR	-2.18	0.49
HAVCR1__LDLRAP1	HAVCR1	LDLRAP1	-0.56	0.84
HAVCR1__LPL	HAVCR1	LPL	-1.65	0.59
HAVCR1__MLXIPL	HAVCR1	MLXIPL	-0.63	1.35
HAVCR1__MYBPHL	HAVCR1	MYBPHL	0.59	1.31
HAVCR1__MYLIP	HAVCR1	MYLIP	-0.13	0.65
HAVCR1__NCAN	HAVCR1	NCAN	-0.51	1.07
HAVCR1__NEG9	HAVCR1	NEG9	0.58	1.18
HAVCR1__PAFAH1B1	HAVCR1	PAFAH1B1	-0.97	0.90
HAVCR1__PAFAH1B2	HAVCR1	PAFAH1B2	0.13	0.30
HAVCR1__PCSK9	HAVCR1	PCSK9	-0.67	1.93
HAVCR1__PVRL2	HAVCR1	PVRL2	-0.70	1.75
HAVCR1__SEZ6L	HAVCR1	SEZ6L	-1.19	0.86
HAVCR1__SIK3	HAVCR1	SIK3	1.23	1.74
HAVCR1__SORT1	HAVCR1	SORT1	-2.09	0.91
HAVCR1__TM6SF2	HAVCR1	TM6SF2	-0.41	1.15
HAVCR1__TMEM57	HAVCR1	TMEM57	-0.96	0.50
HAVCR1__TOMM40	HAVCR1	TOMM40	-1.06	1.55
HAVCR1__WDR12	HAVCR1	WDR12	-0.48	0.71
HAVCR1__ZNF259	HAVCR1	ZNF259	-1.01	0.53
HMGCR__HMGCR	HMGCR	HMGCR	-0.63	0.80
HMGCR__LDLR	HMGCR	LDLR	-0.22	2.48
HMGCR__LDLRAP1	HMGCR	LDLRAP1	1.24	0.57
HMGCR__LPL	HMGCR	LPL	0.11	0.43
HMGCR__MLXIPL	HMGCR	MLXIPL	2.17	2.23
HMGCR__MYBPHL	HMGCR	MYBPHL	1.74	0.51
HMGCR__MYLIP	HMGCR	MYLIP	-0.51	0.75
HMGCR__NCAN	HMGCR	NCAN	1.38	1.22
HMGCR__NEG9	HMGCR	NEG9	0.71	1.11
HMGCR__PAFAH1B1	HMGCR	PAFAH1B1	1.45	1.74
HMGCR__PAFAH1B2	HMGCR	PAFAH1B2	0.00	0.90
HMGCR__PCSK9	HMGCR	PCSK9	0.44	0.80
HMGCR__PVRL2	HMGCR	PVRL2	1.24	1.14
HMGCR__SEZ6L	HMGCR	SEZ6L	1.26	0.39
HMGCR__SIK3	HMGCR	SIK3	1.71	1.38
HMGCR__SORT1	HMGCR	SORT1	1.02	1.44
HMGCR__TM6SF2	HMGCR	TM6SF2	0.76	0.72
HMGCR__TMEM57	HMGCR	TMEM57	0.94	0.97
HMGCR__TOMM40	HMGCR	TOMM40	0.62	1.00
HMGCR__WDR12	HMGCR	WDR12	0.85	0.92
HMGCR__ZNF259	HMGCR	ZNF259	0.11	1.21
LDLRAP1__LDLRAP1	LDLRAP1	LDLRAP1	0.86	0.98
LDLRAP1__LPL	LDLRAP1	LPL	-0.81	0.96
LDLRAP1__MLXIPL	LDLRAP1	MLXIPL	1.50	1.94
LDLRAP1__MYBPHL	LDLRAP1	MYBPHL	0.90	2.03
LDLRAP1__MYLIP	LDLRAP1	MYLIP	0.68	1.88
LDLRAP1__NCAN	LDLRAP1	NCAN	1.85	2.37
LDLRAP1__NEG9	LDLRAP1	NEG9	0.94	1.45

LDLRAP1__PAFAH1B1	LDLRAP1	PAFAH1B1	-0.73	0.94
LDLRAP1__PAFAH1B2	LDLRAP1	PAFAH1B2	0.57	1.24
LDLRAP1__PCSK9	LDLRAP1	PCSK9	0.43	1.92
LDLRAP1__PVRL2	LDLRAP1	PVRL2	0.72	1.81
LDLRAP1__SEZ6L	LDLRAP1	SEZ6L	-1.46	1.36
LDLRAP1__SIK3	LDLRAP1	SIK3	1.51	1.00
LDLRAP1__SORT1	LDLRAP1	SORT1	-1.11	0.77
LDLRAP1__TM6SF2	LDLRAP1	TM6SF2	-0.57	0.49
LDLRAP1__TMEM57	LDLRAP1	TMEM57	0.16	1.36
LDLRAP1__TOMM40	LDLRAP1	TOMM40	-0.96	2.20
LDLRAP1__WDR12	LDLRAP1	WDR12	0.20	1.03
LDLRAP1__ZNF259	LDLRAP1	ZNF259	0.11	1.39
LDLR__LDLR	LDLR	LDLR	-1.39	1.17
LDLR__LDLRAP1	LDLR	LDLRAP1	-2.44	0.65
LDLR__LPL	LDLR	LPL	-2.05	0.65
LDLR__MLXIPL	LDLR	MLXIPL	-2.03	0.98
LDLR__MYBPHL	LDLR	MYBPHL	-1.54	1.87
LDLR__MYLIP	LDLR	MYLIP	-1.73	1.16
LDLR__NCAN	LDLR	NCAN	-1.86	0.69
LDLR__NEG9	LDLR	NEG9	-1.50	1.17
LDLR__PAFAH1B1	LDLR	PAFAH1B1	-1.24	1.74
LDLR__PAFAH1B2	LDLR	PAFAH1B2	-1.44	1.52
LDLR__PCSK9	LDLR	PCSK9	-1.84	0.97
LDLR__PVRL2	LDLR	PVRL2	0.08	3.67
LDLR__SEZ6L	LDLR	SEZ6L	-1.96	0.64
LDLR__SIK3	LDLR	SIK3	-1.61	0.94
LDLR__SORT1	LDLR	SORT1	-0.91	0.53
LDLR__TM6SF2	LDLR	TM6SF2	-1.07	1.76
LDLR__TMEM57	LDLR	TMEM57	-1.98	0.69
LDLR__TOMM40	LDLR	TOMM40	-2.18	1.50
LDLR__WDR12	LDLR	WDR12	-1.69	1.01
LDLR__ZNF259	LDLR	ZNF259	-1.45	1.11
LPL__LPL	LPL	LPL	-1.41	0.88
LPL__MLXIPL	LPL	MLXIPL	0.36	0.76
LPL__MYBPHL	LPL	MYBPHL	-0.35	0.85
LPL__MYLIP	LPL	MYLIP	0.53	1.00
LPL__NCAN	LPL	NCAN	0.66	0.79
LPL__NEG9	LPL	NEG9	-0.29	1.12
LPL__PAFAH1B1	LPL	PAFAH1B1	-1.35	1.44
LPL__PAFAH1B2	LPL	PAFAH1B2	-0.86	1.08
LPL__PCSK9	LPL	PCSK9	-1.23	0.83
LPL__PVRL2	LPL	PVRL2	-0.43	1.24
LPL__SEZ6L	LPL	SEZ6L	-0.79	1.27
LPL__SIK3	LPL	SIK3	0.20	1.03
LPL__SORT1	LPL	SORT1	-1.08	1.50
LPL__TM6SF2	LPL	TM6SF2	-0.82	1.31
LPL__TMEM57	LPL	TMEM57	-0.68	0.88
LPL__TOMM40	LPL	TOMM40	-2.23	0.82
LPL__WDR12	LPL	WDR12	-0.73	1.02
LPL__ZNF259	LPL	ZNF259	-0.71	1.25
MLXIPL__MLXIPL	MLXIPL	MLXIPL	1.54	0.79
MLXIPL__MYBPHL	MLXIPL	MYBPHL	-0.06	1.69
MLXIPL__MYLIP	MLXIPL	MYLIP	1.30	1.13
MLXIPL__NCAN	MLXIPL	NCAN	1.68	0.71
MLXIPL__NEG9	MLXIPL	NEG9	1.21	1.75
MLXIPL__PAFAH1B1	MLXIPL	PAFAH1B1	-0.54	0.81
MLXIPL__PAFAH1B2	MLXIPL	PAFAH1B2	0.32	1.30
MLXIPL__PCSK9	MLXIPL	PCSK9	1.87	1.03
MLXIPL__PVRL2	MLXIPL	PVRL2	2.20	1.93
MLXIPL__SEZ6L	MLXIPL	SEZ6L	-0.91	2.15
MLXIPL__SIK3	MLXIPL	SIK3	0.70	1.05
MLXIPL__SORT1	MLXIPL	SORT1	-0.14	0.75
MLXIPL__TM6SF2	MLXIPL	TM6SF2	-0.60	1.51
MLXIPL__TMEM57	MLXIPL	TMEM57	2.17	0.76
MLXIPL__TOMM40	MLXIPL	TOMM40	-2.14	1.24
MLXIPL__WDR12	MLXIPL	WDR12	-0.23	1.78
MLXIPL__ZNF259	MLXIPL	ZNF259	-0.95	1.18
MYBPHL__MYBPHL	MYBPHL	MYBPHL	0.98	0.24
MYBPHL__MYLIP	MYBPHL	MYLIP	0.72	0.90
MYBPHL__NCAN	MYBPHL	NCAN	1.10	1.58
MYBPHL__NEG9	MYBPHL	NEG9	0.06	1.56
MYBPHL__PAFAH1B1	MYBPHL	PAFAH1B1	0.66	1.92
MYBPHL__PAFAH1B2	MYBPHL	PAFAH1B2	-0.60	0.91
MYBPHL__PCSK9	MYBPHL	PCSK9	-0.31	0.80
MYBPHL__PVRL2	MYBPHL	PVRL2	0.62	1.42
MYBPHL__SEZ6L	MYBPHL	SEZ6L	0.73	1.44
MYBPHL__SIK3	MYBPHL	SIK3	2.17	1.49
MYBPHL__SORT1	MYBPHL	SORT1	1.08	2.25
MYBPHL__TM6SF2	MYBPHL	TM6SF2	0.02	0.74
MYBPHL__TMEM57	MYBPHL	TMEM57	0.25	0.92
MYBPHL__TOMM40	MYBPHL	TOMM40	0.21	0.90
MYBPHL__WDR12	MYBPHL	WDR12	0.92	1.02
MYBPHL__ZNF259	MYBPHL	ZNF259	1.25	1.43
MYLIP__MYLIP	MYLIP	MYLIP	0.54	1.15
MYLIP__NCAN	MYLIP	NCAN	0.82	1.24
MYLIP__NEG9	MYLIP	NEG9	0.38	1.53
MYLIP__PAFAH1B1	MYLIP	PAFAH1B1	-0.10	1.36
MYLIP__PAFAH1B2	MYLIP	PAFAH1B2	-0.17	1.32
MYLIP__PCSK9	MYLIP	PCSK9	1.07	2.68
MYLIP__PVRL2	MYLIP	PVRL2	0.06	0.95
MYLIP__SEZ6L	MYLIP	SEZ6L	2.18	0.97
MYLIP__SIK3	MYLIP	SIK3	0.53	0.76
MYLIP__SORT1	MYLIP	SORT1	0.34	1.19
MYLIP__TM6SF2	MYLIP	TM6SF2	0.94	0.53
MYLIP__TMEM57	MYLIP	TMEM57	1.22	0.66
MYLIP__TOMM40	MYLIP	TOMM40	0.71	1.68
MYLIP__WDR12	MYLIP	WDR12	-0.31	0.29
MYLIP__ZNF259	MYLIP	ZNF259	0.39	1.17

NCAN__NCAN	NCAN	NCAN	1.42	0.59
NCAN__NEG9	NCAN	NEG9	1.04	1.76
NCAN__PAFAH1B1	NCAN	PAFAH1B1	-0.36	0.70
NCAN__PAFAH1B2	NCAN	PAFAH1B2	1.68	1.17
NCAN__PCSK9	NCAN	PCSK9	1.62	2.54
NCAN__PVRL2	NCAN	PVRL2	2.35	1.79
NCAN__SEZ6L	NCAN	SEZ6L	-0.59	1.10
NCAN__SIK3	NCAN	SIK3	1.49	1.43
NCAN__SORT1	NCAN	SORT1	-0.51	0.98
NCAN__TM6SF2	NCAN	TM6SF2	-0.22	0.58
NCAN__TMEM57	NCAN	TMEM57	0.96	2.45
NCAN__TOMM40	NCAN	TOMM40	-1.40	1.58
NCAN__WDR12	NCAN	WDR12	0.35	1.80
NCAN__ZNF259	NCAN	ZNF259	0.01	0.39
NEG9__NEG9	NEG9	NEG9	0.00	1.00
NEG9__NPC1	NEG9	NPC1	0.02	1.14
NEG9__PAFAH1B1	NEG9	PAFAH1B1	-0.55	1.72
NEG9__PAFAH1B2	NEG9	PAFAH1B2	0.07	1.21
NEG9__PCSK9	NEG9	PCSK9	-0.01	1.39
NEG9__PVRL2	NEG9	PVRL2	0.21	0.94
NEG9__SEZ6L	NEG9	SEZ6L	-0.23	1.32
NEG9__SIK3	NEG9	SIK3	0.48	1.75
NEG9__SORT1	NEG9	SORT1	-0.39	0.98
NEG9__TM6SF2	NEG9	TM6SF2	-0.11	1.32
NEG9__TMEM57	NEG9	TMEM57	0.02	1.23
NEG9__TOMM40	NEG9	TOMM40	-0.63	1.73
NEG9__WDR12	NEG9	WDR12	-0.38	1.08
NEG9__ZNF259	NEG9	ZNF259	-0.30	1.22
PAFAH1B1__PAFAH1B1	PAFAH1B1	PAFAH1B1	-1.57	0.82
PAFAH1B1__PAFAH1B2	PAFAH1B1	PAFAH1B2	-0.69	1.14
PAFAH1B1__PCSK9	PAFAH1B1	PCSK9	-1.76	1.79
PAFAH1B1__PVRL2	PAFAH1B1	PVRL2	1.41	1.75
PAFAH1B1__SEZ6L	PAFAH1B1	SEZ6L	0.28	0.18
PAFAH1B1__SIK3	PAFAH1B1	SIK3	1.62	0.48
PAFAH1B1__SORT1	PAFAH1B1	SORT1	0.70	2.60
PAFAH1B1__TM6SF2	PAFAH1B1	TM6SF2	0.08	0.80
PAFAH1B1__TMEM57	PAFAH1B1	TMEM57	0.08	1.54
PAFAH1B1__TOMM40	PAFAH1B1	TOMM40	0.89	0.90
PAFAH1B1__WDR12	PAFAH1B1	WDR12	0.41	2.05
PAFAH1B1__ZNF259	PAFAH1B1	ZNF259	0.53	1.35
PAFAH1B2__PAFAH1B2	PAFAH1B2	PAFAH1B2	0.13	1.36
PAFAH1B2__PCSK9	PAFAH1B2	PCSK9	-0.50	0.84
PAFAH1B2__PVRL2	PAFAH1B2	PVRL2	-0.76	0.72
PAFAH1B2__SEZ6L	PAFAH1B2	SEZ6L	-0.36	0.81
PAFAH1B2__SIK3	PAFAH1B2	SIK3	-0.20	1.87
PAFAH1B2__SORT1	PAFAH1B2	SORT1	0.91	1.65
PAFAH1B2__TM6SF2	PAFAH1B2	TM6SF2	0.19	1.12
PAFAH1B2__TMEM57	PAFAH1B2	TMEM57	-0.30	0.62
PAFAH1B2__TOMM40	PAFAH1B2	TOMM40	-0.62	1.22
PAFAH1B2__WDR12	PAFAH1B2	WDR12	-0.13	0.72
PAFAH1B2__ZNF259	PAFAH1B2	ZNF259	0.37	0.69
PCSK9__PCSK9	PCSK9	PCSK9	-0.64	0.61
PCSK9__PVRL2	PCSK9	PVRL2	0.32	0.97
PCSK9__SEZ6L	PCSK9	SEZ6L	-1.52	1.15
PCSK9__SIK3	PCSK9	SIK3	-0.42	0.87
PCSK9__SORT1	PCSK9	SORT1	0.09	0.60
PCSK9__TM6SF2	PCSK9	TM6SF2	-0.68	1.62
PCSK9__TMEM57	PCSK9	TMEM57	1.46	0.81
PCSK9__TOMM40	PCSK9	TOMM40	-1.66	0.47
PCSK9__WDR12	PCSK9	WDR12	-0.26	0.79
PCSK9__ZNF259	PCSK9	ZNF259	0.12	0.64
PVRL2__PVRL2	PVRL2	PVRL2	0.67	0.79
PVRL2__SEZ6L	PVRL2	SEZ6L	0.36	0.38
PVRL2__SIK3	PVRL2	SIK3	0.29	1.10
PVRL2__SORT1	PVRL2	SORT1	1.36	1.71
PVRL2__TM6SF2	PVRL2	TM6SF2	-0.38	1.22
PVRL2__TMEM57	PVRL2	TMEM57	1.55	1.38
PVRL2__TOMM40	PVRL2	TOMM40	0.43	1.27
PVRL2__WDR12	PVRL2	WDR12	0.71	1.33
PVRL2__ZNF259	PVRL2	ZNF259	1.12	0.85
SEZ6L__SEZ6L	SEZ6L	SEZ6L	-1.54	0.65
SEZ6L__SIK3	SEZ6L	SIK3	-0.02	0.76
SEZ6L__SORT1	SEZ6L	SORT1	-0.75	0.83
SEZ6L__TM6SF2	SEZ6L	TM6SF2	0.42	0.67
SEZ6L__TMEM57	SEZ6L	TMEM57	0.20	1.87
SEZ6L__TOMM40	SEZ6L	TOMM40	-0.19	1.00
SEZ6L__WDR12	SEZ6L	WDR12	-1.53	0.82
SEZ6L__ZNF259	SEZ6L	ZNF259	0.36	0.63
SIK3__SIK3	SIK3	SIK3	2.35	1.24
SIK3__SORT1	SIK3	SORT1	1.00	1.27
SIK3__TM6SF2	SIK3	TM6SF2	-0.85	1.92
SIK3__TMEM57	SIK3	TMEM57	1.38	0.19
SIK3__TOMM40	SIK3	TOMM40	1.41	0.23
SIK3__WDR12	SIK3	WDR12	0.66	1.35
SIK3__ZNF259	SIK3	ZNF259	1.11	2.85
SORT1__SORT1	SORT1	SORT1	0.16	0.14
SORT1__TM6SF2	SORT1	TM6SF2	0.81	0.47
SORT1__TMEM57	SORT1	TMEM57	0.79	0.74
SORT1__TOMM40	SORT1	TOMM40	0.77	1.04
SORT1__WDR12	SORT1	WDR12	-0.32	0.46
SORT1__ZNF259	SORT1	ZNF259	0.35	0.50
TM6SF2__TM6SF2	TM6SF2	TM6SF2	-0.13	0.46
TM6SF2__TMEM57	TM6SF2	TMEM57	1.11	0.41
TM6SF2__TOMM40	TM6SF2	TOMM40	0.42	1.53
TM6SF2__WDR12	TM6SF2	WDR12	-0.90	0.73
TM6SF2__ZNF259	TM6SF2	ZNF259	0.63	0.84
TMEM57__TMEM57	TMEM57	TMEM57	1.53	0.29
TMEM57__TOMM40	TMEM57	TOMM40	-0.83	1.86

TMEM57__WDR12	TMEM57	WDR12	1.20	1.41
TMEM57__ZNF259	TMEM57	ZNF259	0.32	0.68
TOMM40__TOMM40	TOMM40	TOMM40	0.72	1.11
TOMM40__WDR12	TOMM40	WDR12	-0.18	1.96
TOMM40__ZNF259	TOMM40	ZNF259	0.46	0.31
WDR12__WDR12	WDR12	WDR12	-0.86	0.16
WDR12__ZNF259	WDR12	ZNF259	-0.14	0.80
ZNF259__ZNF259	ZNF259	ZNF259	-0.27	0.23

Table S20. Results of the validation experiments for LDL uptake.

		Robust Z score			
	Treatment	**Exp.1**	**Exp.2**	**Exp.3**	**Exp.4**
1	BCAM __ LDLRAP1	-0.97	-2.18	-1.16	
2	CXCL12__ PAFAH1B1	0.29	-1.97	-3.21	
3	HAVCR1 __ LDLRAP1	-1.97	0.53	-1.2	
4	HAVCR1 __ LPL	-1.77	0.81	-0.12	
5	HAVCR1 __ MLXIPL	-2.88	2.12	-0.34	
6	HAVCR1 __ NCAN	1.01	0.71	-0.27	
7	HAVCR1__ SEZ6L	-0.96	1.22	-0.04	-1.74
8	HAVCR1 __ SORT1	-1.25	1.51	-0.45	-1.56
9	HAVCR1 __ TMEM57	0.18	-0.53	0.98	
10	LDLR__ LDLRAP1	-1.05	-0.17	0.11	-3.37
11	LDLRAP1 __ LPL	2.6	-0.34	-1.2	
12	LDLRAP1 __ SEZ6L	-0.74	-0.01	1.04	
13	MLXIPL __ SEZ6L	0.75	-0.73	-0.46	
14	MLXIPL __ TOMM40	-1.91	0.75	-1.29	-2.07
15	PAFAH1B1 __ TOMM40	-0.66	-0.57	-0.35	
16	APOB __ HMGCR	-0.52	-2.7	6.27	
17	APOE __ MLXIPL	-2.37	2.95	0.58	
18	BCAM __ HAVCR1	-3.88	0.11	0.9	
19	CELSR2 __ LPL	-0.35	-1.09	-0.06	
20	CELSR2 __ SIK3	4.48	-0.57	1.5	
21	HAVCR1 __ MYLIP	-0.47	0.72	3.03	
22	HAVCR1 __ ZNF259	-2.07	0.99	-0.26	0.73
23	HMGCR __ PAFAH1B1	-2.03	-5.24	3.38	
24	LDLR __ NCAN	-2.37	0.52	-1.21	
25	LDLR __ MLXIPL	-1.42	0.1	-0.88	
26	LDLR __ PVRL2	-1.05	-1.47	-0.94	
27	LDLRAP1 __ SORT1	-0.43	2.56	-1.7	-1.35
28	MYBPHL __ SIK3	0.33	1.68	0.36	
29	MYLIP __ SEZ6L	2.28	-0.28	-1.47	
30	NCAN __ SEZ6L	-0.48	0.01	-1.55	
31	NCAN __ TOMM40	-0.88	-0.06	-3.12	
32	PAFAH1B1 __ SIK3	-4.04	4.45	3.55	
33	PCSK9 __ TMEM57	1.95	2.02	-0.19	
34	PVRL2 __ SORT1	1.85	0.67	0.8	
35	SORT1 __ TOMM40	3.28	0.68	0.87	

Table S21. The 144 SNPs that were tested for SNP co-occurrence in data from the Bioimage study.

SNPs that were tested for SNP co-occurrence in data from the Bioimage study are listed. The 133 SNPs were extracted from the PhenoScanner of the University of Cambridge, with lipid-trait associations from the study of Willer et al. 2013. Associations for other traits (CAD, CHD, MI) were extracted from other studies and are highlighted in green. Shown are the chromosomal locus where the SNP is located (Locus), the Reference SNP cluster ID (rs ID), the SNP position on the chromosome (Position), the risk (A1) and the ancestral (A2) allele, the SNP function -if known- (Function), the reported gene (Gene), the associated trait (Trait), the p-value for the association (p-value), the number of individuals tested (N), the reported odds ratio or beta-coefficient associated with the strongest SNP risk allele (beta) and the standard error of the beta-coefficient (sebeta).

Locus	SNP	Pos	A1	A2	Function	Gene	Trait	p-value	N	beta	Sebeta	Reference
1p13.3	rs629301	109818306	G	T	UTR-3'	CELSR2	LDL	5E-241	NA	-0.167	0	Willer et al. 2013
							TC	2E-170	155873	0.134	0.0047	Willer et al. 2013
	rs599839	109822166	A	G	nearGene-3'	PSRC1	LDL	2.70E-268	170167	0.1601	0.0044	Willer et al. 2013
							TC	1.40E-186	184335	0.1281	0.0042	Willer et al. 2013
	rs12740374	109817590	G	T	UTR-3'	CELSR2	LDL	2.4e-272	172820	0.161	0.0044	Willer et al. 2013
							TC	7.70E-187	187071	0.1278	0.0043	Willer et al. 2013
	rs660240	109817838	T	C	UTR-3'	CELSR2	LDL	9.00E-265	173022	-0.1607	0.0044	Willer et al. 2013
							TC	7.00E-182	187288	-0.1276	0.0043	Willer et al. 2013
	rs646776	109818530	T	C	nearGene-3	CELSR2	LDL	1.60E-272	173021	0.1602	0.0044	Willer et al. 2013
							TC	4.80E-187	187288	0.1272	0.0042	Willer et al. 2013
	rs7528419	109817192	A	G	UTR-3'		LDL	1.50E-165	114743	0.1547	0.0054	Willer et al. 2013
							TC	5.60E-110	128033	0.1189	0.0051	Willer et al. 2013
	rs602633	109821511	T	G		CELSR2	LDL	1.50E-261	171593	-0.1591	0.0044	Willer et al. 2013
							TC	2.10E-180	185817	-0.1268	0.0043	Willer et al. 2013
	rs4970833	109804646	A	G	intron	CELSR2	LDL	7.00E-55	160040	0.0637	0.0039	Willer et al. 2013
							TC	9.70E-38	171955	0.0503	0.0038	Willer et al. 2013
	rs6689614	109807099	A	G	synonymous	CELSR2	LDL	7.80E-61	172988	0.0646	0.0038	Willer et al. 2013
							TC	2.80E-41	187253	0.0507	0.0036	Willer et al. 2013
	rs4970834	109814880	T	C	intron	CELSR2	LDL	2.00E-208	172825	0.1503	0.0047	Willer et al. 2013
							TC	9.60E-143	187081	0.1188	0.0045	Willer et al. 2013
	rs611917	109815252	G	A	intron	CELSR2	LDL	1.90E-151	159987	-0.111	0.0041	Willer et al. 2013
							TC	5.60E-104	171876	-0.0881	0.0039	Willer et al. 2013
	rs3902354	109819296	C	A		CELSR2	CHD	2.1E-11	184305	-0.06963	0.0104	Nikpay et al. 2015
							MI	0.00000014	167181	-0.06024	0.01143	Nikpay et al. 2015
	rs672569	109827253	A	G	nearGene-5'	CELSR2	LDL	2.10E-64	89876	-0.1431	0.0082	Willer et al. 2013
1p32.3	rs2479409	55504650	G	A	nearGene-5'	PCSK9	LDL	3E-50	172970	-0.0642	0.0041	Willer et al. 2013
							TC	2E-39	187226	-0.054	0.004	Willer et al. 2013
	rs11206510	55496039	C	T	intergenic	PCSK9	LDL	2.40E-53	172812	-0.0831	0.005	Willer et al. 2013
							TC	1.10E-41	187066	-0.069	0.0048	Willer et al. 2013
	rs11591147	55505647	G	T	nearGene-5'	PCSK9	LDL	8.60E-143	77417	0.497	0.018	Willer et al. 2013
							TC	8.80E-86	85729	0.3341	0.0173	Willer et al. 2013
	rs499883	55519174	A	G			CHD	0.007	184305	0.02872	0.01066	Nikpay et al. 2015
			A	G			MI	0.034	167181	0.02497	0.01181	Nikpay et al. 2015
	rs505151	55529187	A	G	ncRNA	PCSK9	LDL	4.2e-17	151827	-0.0866	0.01	Willer et al. 2013
							TC	3.8e-10	164218	-0.0619	0.0096	Willer et al. 2013
1p36.11	rs12027135	25449242	A	T	intron	TMEM57	TC	5E-12	NA			Willer et al. 2013
							LDL	2E-14	NA			Willer et al. 2013
2p24.1	rs1367117	21263900	A	G	missense	APOB	LDL	9.48E-183	173007	0.1186	0.004	Willer et al. 2013
							TC	2.48E-139	187252	0.0995	0.0038	Willer et al. 2013
	rs10199768	21244000	G	T	intron	APOB	LDL	5.00E-139	170875	-0.0971	0.0037	Willer et al. 2013
							TC	2.80E-108	185137	-0.083	0.0036	Willer et al. 2013
							TG	5.70E-21	175642	-0.0353	0.0035	Willer et al. 2013
	rs693	21232195	G	A	cds-synon	APOB	LDL	1.20E-131	157125	-0.0954	0.0038	Willer et al. 2013
							TC	5.20E-102	168987	-0.0816	0.0037	Willer et al. 2013
							TG	2.20E-30	160959	-0.0415	0.0034	Willer et al. 2013
	rs1801701	21228827	T	C	missense	APOB	LDL	2.3e-21	173030	0.0638	0.0064	Willer et al. 2013
							TC	8.3e-15	187293	0.0497	0.0062	Willer et al. 2013
	rs1042034	21225281	T	C	missense	APOB	TG	4.10E-66	177771	0.0704	0.0039	Willer et al. 2013
							HDL	2.90E-54	187071	-0.0659	0.004	Willer et al. 2013
							LDL	7.40E-39	173002	0.0586	0.0043	Willer et al. 2013
							TC	2.50E-25	187265	0.0448	0.0042	Willer et al. 2013
	rs12713956	21241505	G	A	intron	APOB	LDL	1.3e-28	173010	-0.0719	0.0062	Willer et al. 2013
							TC	2.9e-28	187270	-0.0688	0.006	Willer et al. 2013
	rs11902417	21198900	A	G	intergenic	APOB	TG	6.70E-67	173384	-0.0694	0.0039	Willer et al. 2013
							HDL	1.20E-47	182623	0.0605	0.004	Willer et al. 2013
							LDL	2.10E-46	168630	-0.0628	0.0043	Willer et al. 2013
							TC	4.90E-33	182782	-0.0506	0.0041	Willer et al. 2013
	rs515135	21286057	T	C	intergenic	APOB	LDL	1.10E-178	173029	-0.1394	0.0048	Willer et al. 2013
							TC	6.40E-151	187291	-0.1238	0.0046	Willer et al. 2013
	rs6544366	21204025	G	T	intergenic	APOB	TG	5.40E-66	174737	0.0685	0.0039	Willer et al. 2013
							HDL	1.80E-48	184086	-0.0606	0.004	Willer et al. 2013
							LDLc	9.30E-46	170058	0.062	0.0043	Willer et al. 2013
							TC	8.70E-32	184227	0.0492	0.0041	Willer et al. 2013
	rs6754295	21206183	G	T	intergenic	APOB	TG	2.00E-62	177784	-0.0657	0.0038	Willer et al. 2013
							HDL	1.20E-49	187085	0.0609	0.0039	Willer et al. 2013
							LDL	1.60E-47	173013	-0.0628	0.0042	Willer et al. 2013
							TC	2.90E-33	187280	-0.0501	0.0041	Willer et al. 2013

	rs7557067	21208211	A	G	intergenic	APOB	Tg	2.00E-62	177772	0.0659	0.0038	Willer et al. 2013
							HDL	2.20E-49	187078	-0.0608	0.0039	Willer et al. 2013
							LDL	2.50E-46	173006	0.062	0.0042	Willer et al. 2013
							TC	1.20E-32	187265	0.0495	0.0041	Willer et al. 2013
	rs676210	21231524	A	G	missense	MYLIP	Tg	3.30E-71	177782	-0.0733	0.0039	Willer et al. 2013
							HDL	2.30E-54	187081	0.066	0.004	Willer et al. 2013
	rs934197	21267461	A	G	non-coding		LDL	1.60E-74	84950	0.1049	0.0054	Willer et al. 2013
							HDL	3.30E-05	3.30E-05	-0.0228	0.0051	Willer et al. 2013
	rs673548	21237544	A	G	intron	APOB	TG	2.00E-68	163871	-0.0742	0.0041	Willer et al. 2013
							HDL	2.10E-53	172585	0.0678	0.0042	Willer et al. 2013
							LDL	2.70E-38	159959	-0.0606	0.0045	Willer et al. 2013
	rs754523	21311691	G	A	intergenic	APOB	LDL	6.00E-122	160051	0.0981	0.004	Willer et al. 2013
							TC	2.60E-91	171993	0.0819	0.0039	Willer et al. 2013
	rs1713222	21271323	G	A	intergenic	APOB	LDL	2.60E-130	160080	0.1327	0.0053	Willer et al. 2013
							TC	1.70E-113	172022	0.121	0.0051	Willer et al. 2013
	rs4635554	21389659	G	T		APOB	LDL	1.20E-75	156927	0.0783	0.0041	Willer et al. 2013
							TC	5.10E-63	169635	0.0691	0.004	Willer et al. 2013
	rs312985	21378805	G	A		APOB	LDL	1.40E-118	158796	0.1106	0.0046	Willer et al. 2013
							TC	7.80E-102	170715	0.0988	0.0045	Willer et al. 2013
	rs506585	21397182	A	G		APOB	LDL	3.90E-119	160084	0.1099	0.0046	Willer et al. 2013
							TC	7.80E-102	172027	0.0983	0.0045	Willer et al. 2013
	rs503662	21414142	T	C		APOB	LDL	3.80E-83	173020	-0.0836	0.0042	Willer et al. 2013
							TC	2.60E-70	187287	-0.0739	0.0041	Willer et al. 2013
	rs531819	21263639	G	T	intron	APOB	LDL	4.00E-141	173033	0.1343	0.0052	Willer et al. 2013
							TC	6.50E-123	187299	0.1222	0.005	Willer et al. 2013
	rs17398765	21270751	A	G		APOB	LDL	3.50E-32	168107	-0.0916	0.0076	Willer et al. 2013
							TC	9.90E-24	182233	-0.0734	0.0073	Willer et al. 2013
	rs7582720	202881162	T	C	intron	WDR12	LDL	0.000025	170099	-0.0249	0.0055	Willer et al. 2013
							CHD	7.5E-18	184305	0.13342	0.0155	Nikpay et al. 2015
2q33.2							MI	1.7E-13	167181	0.12394	0.01681	Nikpay et al. 2015
	rs6725887			C	intron	WDR12	CHD	9.5E-18	184305	0.13288	0.01549	Nikpay et al. 2015
							MI	2.6E-12	167181	0.11885	0.01699	Nikpay et al. 2015
	rs12916	75360714	T	C	UTR-3'	HMGCR	TC	5E-74	182530	0.0684	0.0036	Willer et al. 2013
							LDL	8E-78	NA	0.073	0	Willer et al. 2013
	rs3846662	74651084	G	A	intron	HMGCR	LDL	2.20E-69	NA			Willer et al. 2013
							TC	6.90E-66	NA			Willer et al. 2013
	rs3846663	74655726	C	T	intron	HMGCR	LDL	1.10E-75	170020	-0.0722	0.0037	Willer et al. 2013
							TC	2.90E-70	184190	-0.0665	0.0036	Willer et al. 2013
5q13.3	**rs12654264**	74648603	A	T	intron	HMGCR		3.90E-70	172958	-0.0687	0.0037	Willer et al. 2013
								7.80E-65	187215	-0.0632	0.0036	Willer et al. 2013
	rs7703051	74625487	A	C	intergenic	HMGCR	LDL	1.40E-77	173015	0.0727	0.0037	Willer et al. 2013
							TC	3.10E-72	187283	0.0672	0.0036	Willer et al. 2013
	rs10045497	74636484	A	C	intron	HMGCR	TC	2.10E-46	94595	0.0768	0.0052	Willer et al. 2013
							LDL	2.30E-44	89888	0.0766	0.0053	Willer et al. 2013
5q21	rs383830				intergenic	FAM174A	CAD	1.3E-5				WTTC 2007
	rs6882076	156390297	C	T	nearGene-5'	TIMD4	TC	5.35E-41	187270	0.0508	0.0037	Willer et al. 2013
							TG	1.51E-15	177778	0.0286	0.0035	Willer et al. 2013
							LDL	3.31E-31	173006	0.0456	0.0038	Willer et al. 2013
5q33.3							HDL	0.6852	187080	0.0015	0.0035	Willer et al. 2013
	rs1501908	156398169	G	C	intergenic	HAVCR1	TC	4.9E-39	187107	-0.0494	0.0037	Willer et al. 2013
							LDL	1.1E-28	172869	-0.0436	0.0038	Willer et al. 2013
							TG	8.3E-15	177635	-0.0278	0.0035	Willer et al. 2013
6p22.3	**rs3757354**	16127176	C	T	nearGene-5'	MYLIP	LDL	2E-17	172987	-0.0382	0.0044	Willer et al. 2013
							TC	2E-15	187247	-0.0348	0.0042	Willer et al. 2013
	rs17145738	72982874	C	T	nearGene-3'	TBL2	TG	9.42E-99	175664	0.1149	0.0053	Willer et al. 2013
							HDL	4.95E-13	184970	0.0408	0.0053	Willer et al. 2013
							LDL	0.5431	170891	0.0039	0.0057	Willer et al. 2013
							TC	0.008813	185161	0.0136	0.0055	Willer et al. 2013
	rs17145750	73612048	C	T	intron	MLXIPL	HDL	5E-13	NA			Willer et al. 2013
	rs2074755	72877166	T	C	intron	BAZ1B	TG	1.4E-98	174001	0.1092	0.0051	Willer et al. 2013
							HDL	3.7E-12	183352	-0.0393	0.0052	Willer et al. 2013
	rs2286276	72987354	T	C	intron	TBL2	TG	3.7e-70	163993	-0.067	0.0038	Willer et al. 2013
							HDL	1.3e-06	172719	0.0208	0.0039	Willer et al. 2013
	rs2240466	72856269	G	A	UTR-3'	BAZ1B	TG	5.30E-90	170671	0.1051	0.0052	Willer et al. 2013
							HDL	1.00E-11	178896	-0.0394	0.0053	Willer et al. 2013
7q11.23	**rs1178979**	72856430	T	A	UTR-3'	BAZ1B	TG	2.00E-97	177782	0.0895	0.0042	Willer et al. 2013
	rs714052	72864869	A	T	intron	BAZ1B	TG	2.00E-99	177802	-0.1084	0.005	Willer et al. 2013
							HDL	2.20E-12	187107	0.0394	0.0052	Willer et al. 2013
	rs3812316	73020337	C	G	missense	MLXIPL	TG	1.30E-31	79868	0.0912	0.0074	Willer et al. 2013
							HDL	1.50E-04	84783	-0.0317	0.0077	Willer et al. 2013
	rs17145750	73026378	C	T	intron	MLXIPL	TG	1.00E-99	177797	0.1015	0.0047	Willer et al. 2013
							HDL	5.70E-12	187102	-0.0358	0.0049	Willer et al. 2013
	rs12056034	72878645	A	G	intron	BAZ1B	TG	9.30E-100	177044	0.109	0.0051	Willer et al. 2013
							HDL	1.30E-12	186350	-0.0399	0.0052	Willer et al. 2013
	rs799160	73060006	T	C	intergenic	MLXIPL	TG	5.50E-30	162118	0.0398	0.0036	Willer et al. 2013
							HDL	2.90E-04	171614	-0.0133	0.0037	Willer et al. 2013
	rs11974409	72989390	A	G	intron	TBL2	TG	1.40E-100	177786	0.0899	0.0042	Willer et al. 2013
	rs12678919	19844222	G	A	intergenic	LPL	TG	1.82E-199	177749	0.1702	0.0056	Willer et al. 2013
							HDL	1.38E-149	187049	0.1554	0.0057	Willer et al. 2013
8p21.3							LDL	0.5045	172984	0.008	0.0061	Willer et al. 2013
							TC	0.7352	187233	0.0003	0.0059	Willer et al. 2013

	rs17411031	19852310	C	G	intergenic	LPL	TG	7.30E-180	177569	0.1075	0.0037	Willer et al. 2013
							HDL	3.90E-150	186853	-0.1047	0.0038	Willer et al. 2013
	rs17482753	19832646	G	T	intergenic	LPL	TG	5.80E-193	177737	0.1654	0.0055	Willer et al. 2013
							HDL	2.80E-149	187036	-0.1538	0.0056	Willer et al. 2013
	rs3289	19823192	C	T	UTR-3'		HDL	6.40E-46	178804	-0.1695	0.0113	Willer et al. 2013
							TG	3.70E-33	169504	0.1447	0.0111	Willer et al. 2013
	rs326	19819439	A	G	intron	LPL	TG	1.00E-63	96966	0.0869	0.005	Willer et al. 2013
							HDL	2.00E-58	102961	-0.0889	0.0051	Willer et al. 2013
	rs10503669	19847690	A	C		LPL	TG	7.70E-190	177722	-0.1676	0.0056	Willer et al. 2013
							HDL	4.10E-144	187021	0.1534	0.0057	Willer et al. 2013
	rs13702	19824492	T	C	UTR-3'	LPL	TG	6.60E-187	177749	0.1069	0.0037	Willer et al. 2013
							HDL	1.30E-160	187044	-0.1058	0.0038	Willer et al. 2013
	rs325	19819328	C	T	intron	LPL	TG	9.40E-195	177748	-0.1676	0.0055	Willer et al. 2013
							HDL	2.10E-148	187045	0.1552	0.0057	Willer et al. 2013
	rs10105606	19827848	A	C	intergenic	LPL	TG	5.10E-133	177734	-0.0876	0.0035	Willer et al. 2013
							HDL	2.70E-105	187034	0.0833	0.0036	Willer et al. 2013
	rs2083637	19865175	G	A	intergenic	LPL	TG	2.10E-175	163987	-0.1084	0.0038	Willer et al. 2013
							HDL	3.80E-143	172714	0.1051	0.0039	Willer et al. 2013
	rs10096633	19830921	C	T	intergenic	LPL	TG	3.40E-182	177749	0.1471	0.005	Willer et al. 2013
							HDL	1.70E-144	187049	-0.1384	0.0051	Willer et al. 2013
	rs328	19819724	C	G	coding-STOP-GAIN	LPL	TG	2.00E-179	160341	0.167	0.0058	Willer et al. 2013
							HDL	3.20E-135	169631	-0.1541	0.0059	Willer et al. 2013
	rs3916027	19824868	G	A	nearGene-3'		TG	1.20E-179	177708	0.1072	0.0037	Willer et al. 2013
							HDL	2.30E-154	187006	-0.1061	0.0038	Willer et al. 2013
	rs327	19819536	G	T	intron	LPL	TG	5.70E-174	177713	-0.1045	0.0037	Willer et al. 2013
							HDL	4.60E-145	187014	0.1024	0.0038	Willer et al. 2013
	rs1581675	19858499	A	T		LPL	TG	7.40E-163	169466	-0.1054	0.0039	Willer et al. 2013
							HDL	2.20E-133	177687	0.1026	0.004	Willer et al. 2013
	rs6993414	19902918	A	G		LPL	TG	3.30E-187	176745	0.1661	0.0056	Willer et al. 2013
							HDL	5.00E-140	186043	-0.1514	0.0057	Willer et al. 2013
	rs2197089	19826373	G	A	intergenic	LPL	TG	7.60E-73	173751	0.0608	0.0034	Willer et al. 2013
							HDL	9.60E-66	182963	-0.0612	0.0035	Willer et al. 2013
	rs6586891	19914598	A	C	intergenic	LPL	TG	9.20E-66	146609	-0.0656	0.0038	Willer et al. 2013
							HDL	4.10E-61	154466	0.0675	0.0039	Willer et al. 2013
	rs331	19820405	A	G	intron	LPL	TG	1.10E-174	172997	-0.1089	0.0039	Willer et al. 2013
							HDL	3.90E-150	182314	0.1105	0.004	Willer et al. 2013
	rs17091905	19849757	G	A		LPL	TG	2.00E-174	177742	-0.1481	0.0051	Willer et al. 2013
							HDL	1.60E-137	187041	0.1377	0.0052	Willer et al. 2013
	rs264	19813180	A	G	intron	LPL	CAD	3E-7				Dichgans et al. 2013
							TG	2.30E-84	172989	-0.0933	0.0048	Willer et al. 2013
							HDL	8.0e-77	182301	0.0976	0.005	Willer et al. 2013
	rs264	19813180	A	G	intron	LPL	HDL	8.0e-77	182301	0.0976	0.005	Willer et al. 2013
							TG	2.30E-84	172989	-0.0933	0.0048	Willer et al. 2013
	rs7016880	19876746	C	G			TG	1.10E-184	177670	-0.165	0.0056	Willer et al. 2013
							HDL	8.10E-144	186956	0.1534	0.0057	Willer et al. 2013
	rs9644568	19928582	A	G		LPL	TG	2.30E-129	177760	-0.1299	0.0053	Willer et al. 2013
							HDL	4.10E-110	187059	0.1266	0.0054	Willer et al. 2013
	rs1059611	19824563	C	T	UTR-3	LPL	TG	1.10E-190	177751	-0.1614	0.0053	Willer et al. 2013
							HDL	1.10E-144	187052	0.1485	0.0055	Willer et al. 2013
	rs79236614	19860460	C	G		LPL	TG	1.10E-79	80066	0.1601	0.0082	Willer et al. 2013
							HDL	4.70E-56	84992	-0.1444	0.0085	Willer et al. 2013
	rs17410962	19848080	G	A		LPL	TG	2.30E-172	176101	-0.1478	0.0051	Willer et al. 2013
							HDL	8.30E-137	185404	0.1379	0.0052	Willer et al. 2013
9p21.3	**rs501120**					CXCL12	CHD	1.40E-11	184305	0.07861	0.01163	Nikpay et al. 2015
							MI	4.10E-08	167181	0.07086	0.01292	Nikpay et al. 2015
	rs1333049	22125503	G			CXCL12	CHD	3.90E-93	184305	0.19322	0.00944	CARDIoGRAMplusC4D
							MI	6.8E-6				Samani 2007
10q11.21	**rs1746048**		C		intergenic	CXCL12	CAD	3E-10				Schunkert 2011
							MI	7E-9				Kathiresan et al. 2009
							CHD	6.50E-11	184305	0.07651	0.01171	Nikpay et al. 2015
10q11.2	**rs266089**				intron	CXCL12	MI	5.2e-18				Yamada 2011
11q23.3	**rs964184**				intron-3'UTR	ZNF259	TG	7e-224	NA	0.0069	90991	Willer et al. 2013
	rs12286037	116652207	C	T	intron	ZNF259	TG	5.30E-193	174655	-0.2151	0.0069	Willer et al. 2013
							HDL	2.80E-47	184011	0.1052	0.007	Willer et al. 2013
	rs3741298	116657561	T	C	intron	ZNF259	TG	2.50E-124	96989	-0.1473	0.0056	Willer et al. 2013
							TC	7.70E-30	102894	-0.067	0.0058	Willer et al. 2013
	rs651821	116662579	C	T	5' UTR	ZNF259	Tg	3.00E-153	116795	0.2358	0.0084	Willer et al. 2013
							TC	5.30E-28	126140	0.0978	0.0089	Willer et al. 2013
							HDL	7.70E-26	126213	-0.1041	0.0088	Willer et al. 2013
	rs2075290	116653296	C	T	intron	APOB	TG	4.80E-244	177775	0.2278	0.0065	Willer et al. 2013
							TC	3.80E-51	187268	0.1056	0.0069	Willer et al. 2013
							HDL	1.30E-39	187082	-0.0947	0.0066	Willer et al. 2013
	rs1558861	116607437	C	T	intergenic	MLXIPL	TG	3.10E-161	103079	0.2368	0.0083	Willer et al. 2013
							TC	1.20E-28	107665	0.1011	0.0088	Willer et al. 2013
							HDL	9.00E-28	109540	-0.0964	0.0083	Willer et al. 2013
	rs4938303	116584987	C	T	intergenic	ZNF259	TG	3.40E-166	177681	0.1071	0.0037	Willer et al. 2013
							HDL	1.90E-31	186981	-0.0462	0.0038	Willer et al. 2013
							TC	5.60E-25	187160	0.0414	0.0039	Willer et al. 2013
	rs28927680	116619073	G	C	UTR-3'	ZNF259	TG	5.50E-78	101632	0.1802	0.0087	Willer et al. 2013
							HDL	2.80E-18	107501	-0.0808	0.0087	Willer et al. 2013

							TC	4.80E-20	107532	0.0883	0.0092	Willer et al. 2013
							LDL	1.20E-09	97617	0.0626	0.0097	Willer et al. 2013
	rs662799	116663707	A	G	intergenic	APOA5	TG	1.60E-241	172713	-0.2474	0.0071	Willer et al. 2013
							TC	2.50E-49	182202	-0.1113	0.0075	Willer et al. 2013
							HDL	4.20E-37	182018	0.1001	0.0072	Willer et al. 2013
							LDL	1.70E-22	168088	-0.0801	0.0078	Willer et al. 2013
	rs12272004	116603724	A	C	intergenic	BUD13	TG	1.60E-179	162406	0.2021	0.0067	Willer et al. 2013
							HDL	1.20E-45	171089	-0.102	0.007	Willer et al. 2013
							TC	1.80E-36	170075	0.0925	0.0072	Willer et al. 2013
							LDL	8.20E-19	158514	0.0679	0.0075	Willer et al. 2013
	rs6589566	116652423	A	G	intron	ZNF259	TG	1.00E-239	177777	-0.226	0.0065	Willer et al. 2013
							HDL	2.70E-38	187084	0.0931	0.0066	Willer et al. 2013
							TC	8.20E-50	187270	-0.1042	0.0069	Willer et al. 2013
							LDL	1.70E-23	173011	-0.0754	0.0071	Willer et al. 2013
	rs2000571	116585533	G	A	intergenic	ZNF259	TG	1.40E-69	177725	-0.0763	0.0041	Willer et al. 2013
							HDL	1.50E-10	187033	0.0291	0.0042	Willer et al. 2013
							TC	2.60E-08	187215	-0.0251	0.0043	Willer et al. 2013
	rs486394	116526322	A	C	intergenic	ZNF259	TG	2.80E-60	177648	-0.0665	0.0038	Willer et al. 2013
							TC	2.70E-14	187138	-0.0307	0.004	Willer et al. 2013
							HDL	2.10E-10	186954	0.0266	0.0038	Willer et al. 2013
							LDL	3.20E-06	172883	-0.0198	0.0042	Willer et al. 2013
	rs2075292	116732512	G	T	intron	SIK3	TG	6.40E-86	177762	0.1026	0.005	Willer et al. 2013
							TC	4.10E-23	187249	0.0534	0.0053	Willer et al. 2013
	rs7124741	116752219	A	T	intron	SIK3	TG	1.30E-85	177760	0.1048	0.0051	Willer et al. 2013
							TC	8.60E-21	187253	0.0521	0.0055	Willer et al. 2013
							LDL	9.10E-06	172995	0.031	0.0057	Willer et al. 2013
	rs17120139	116774201	A	G	intron	SIK3	TG	3.20E-44	91013	0.1045	0.0074	Willer et al. 2013
							TC	2.90E-14	94595	0.0629	0.0082	Willer et al. 2013
							LDL	2.10E-05	89888	0.0367	0.0083	Willer et al. 2013
	rs3135506	116662407	C	G	coding	APOA5	TG	1.10E-68	86746	0.1837	0.0096	Willer et al. 2013
	rs618923	116654159	A	G	intron	ZNF259	TG	9.60E-30	155942	0.0488	0.0041	Willer et al. 2013
							HDL	2.70E-21	165188	-0.0412	0.0042	Willer et al. 2013
	rs139961185	116807343	A	G	intron	SIK3	CHD	2.20E-01	184305	0.07254	0.05942	Nikpay et al. 2015
	rs11216230	116884789	A	G	intron	SIK3	CHD	6.30E-01	184305	-0.01722	0.03592	Nikpay et al. 2015
							MI	8.70E-01	167181	-0.0068	0.04189	Nikpay et al. 2015
17p13.3	**rs9891572**				intergenic	PAFAH1B1	CAD	2.3E-7				Lettre 2011
	rs10401969	19407718	C	T	intron	NCAN	TC	4.13E-77	185666	0.1369	0.007	Willer et al. 2013
							TG	9.70E-70	176172	0.121	0.0065	Willer et al. 2013
							LDLc	2.65E-54	171476	0.1184	0.0072	Willer et al. 2013
							HDL	0.1022	185513	0.0128	0.0068	Willer et al. 2013
	rs2304130	19789528	A	G	intron	NCAN	TC	1.40E-46	170535	0.1039	0.0069	Willer et al. 2013
							TG	5.40E-39	162508	0.085	0.0064	Willer et al. 2013
19p13.11							LDL	2.20E-32	158662	0.0885	0.0072	Willer et al. 2013
	rs17216525	19662220	C	T	intergenic	NCAN	TC	6.70E-62	185680	0.1133	0.0065	Willer et al. 2013
							TG	1.00E-60	176188	0.1035	0.0061	Willer et al. 2013
							LDL	3.40E-43	171492	0.096	0.0067	Willer et al. 2013
	rs16996148	19658472	G	T	intergenic	NCAN	TC	1.50E-62	183979	0.114	0.0065	Willer et al. 2013
							TG	3.20E-60	174491	0.1031	0.0061	Willer et al. 2013
							LDL	2.00E-45	169813	0.0986	0.0067	Willer et al. 2013
	rs6511720	11202306	G	T	intron	LDLR	LDL	3.85E-262	170607	0.2209	0.0061	Willer et al. 2013
							TC	5.43E-202	184763	0.1851	0.0059	Willer et al. 2013
							LDL	3.85E-262	170607	0.2209	0.0061	Willer et al. 2013
							HDL	6.32E-05	184617	0.0249	0.0057	Willer et al. 2013
							TG	0.1043	175280	0.0084	0.0056	Willer et al. 2013
	rs688	11227602	C	T	cds-syn	APOE	LDLc	1.00E-43	166792	-0.054	0.0037	Willer et al. 2013
							TC	1.60E-28	180878	-0.0416	0.0036	Willer et al. 2013
	rs2738459	11238473	A	C	intron	LDLR	LDLc	2.30E-19	88433	0.0532	0.0058	Willer et al. 2013
19p13.2							TC	2.10E-11	93067	0.0387	0.0057	Willer et al. 2013
	rs2228671	11210912	C	T	cds-syn	LDLR	LDL	1.70E-171	170510	0.1768	0.006	Willer et al. 2013
							TC	4.30E-132	184667	0.1475	0.0058	Willer et al. 2013
	rs1122608	11163601	G	T	intron	LDLR	LDL	8.50E-57	159918	0.074	0.0045	Willer et al. 2013
							TC	6.80E-41	173300	0.0599	0.0043	Willer et al. 2013
	rs11668477	11195030	A	G	intergenic	LDLR	LDL	3.80E-143	158659	0.1389	0.0052	Willer et al. 2013
							TC	1.40E-107	170530	0.1144	0.005	Willer et al. 2013
	rs2738446	11227326	C	G	intron	LDLR	LDL	3.30E-29	88417	-0.0621	0.0053	Willer et al. 2013
							TC	3.60E-19	93051	-0.0488	0.0052	Willer et al. 2013
	rs4420638	45422946	G	A	nearGene-3'	APOC1	LDL	2E-178	93103	-0.2251	0.0077	Willer et al. 2013
			G				TC	1E-149	NA	-0.197	0	Willer et al. 2013
			G				HDL	2E-21				Willer et al. 2013
	rs157580	45395266	A	G	intron	TOMM40	LDL	9.20E-119	139565	0.1078	0.0045	Willer et al. 2013
							TC	3.20E-104	151145	0.0969	0.0043	Willer et al. 2013
	rs7412	45412079	C	T	missense	APOE	LDL	0.00E+00	82533	0.5898	0.0101	Willer et al. 2013
19q13.32							TC	1.60E-283	92046	0.3736	0.0096	Willer et al. 2013
	rs519113	45376284	G	C	intron	PVRL2	LDL	1.60E-49	88346	-0.0971	0.0066	Willer et al. 2013
							TC	2.70E-32	92971	-0.0922	0.0066	Willer et al. 2013
	rs439401	45414451	C	T	intergenic	APOE	TG	1.40E-66	152584	0.0659	0.0038	Willer et al. 2013
	rs2075650	45395619	A	G	intron	TOMM40	LDL	1.70E-214	161076	-0.1767	0.0055	Willer et al. 2013
							TC	8.90E-158	175566	-0.1432	0.0052	Willer et al. 2013
	rs769449	45410002	A	G	intron	APOE	LDL	4.90E-161	83660	0.2141	0.0076	Willer et al. 2013
							TC	6.80E-103	93260	0.1589	0.0072	Willer et al. 2013

	rs445925	45415640	G	A	nearGene-5'	APOE	LDL	0.00E+00	105162	0.3634	0.0081	Willer et al. 2013
							TC	2.20E-166	118293	0.2212	0.0077	Willer et al. 2013
	rs8106922	45401666	A	G	intron	APOE	LDL	3.00E-22	139174	-0.0407	0.0041	Willer et al. 2013
	rs405509	45408836	T	G	nearGene-3'	APOE	LDL	2.80E-84	139202	0.0813	0.004	Willer et al. 2013
							TC	1.60E-31	152735	0.0461	0.0039	Willer et al. 2013
	rs769450	45410444	G	A	intron	APOE	LDL	2.40E-11	84849	0.0368	0.0051	Willer et al. 2013
							TC	4.90E-07	94482	0.0267	0.0049	Willer et al. 2013
	rs12721109	45447221	G	A	intron	APOE	LDL	3.00E-122	99409	-0.4462	0.0183	Willer et al. 2013
							TC	8.80E-67	110455	-0.3234	0.0179	Willer et al. 2013
	rs10402271	45329214			intergenic	BCAM	LDL	2.60E-118	171428	0.0927	0.0038	Willer et al. 2013
							TC	8.90E-74	185609	0.0702	0.0037	Willer et al. 2013
	rs4803750	45247627	A	G	intergenic	APOE	LDL	1.70E-162	164086	0.2189	0.0078	Willer et al. 2013
							TC	7.90E-84	178198	0.1485	0.0075	Willer et al. 2013
	rs4605275	45338493	C	T	intergenic	APOE	LDL	2.80E-33	89503	0.0722	0.0061	Willer et al. 2013
							TC	2.60E-20	94142	0.0547	0.006	Willer et al. 2013
	rs16979595	45477381	A	G	intron	APOE	TC	5.40E-09	185650	0.0296	0.005	Willer et al. 2013
							LDL	4.90E-05	171472	0.02	0.0052	Willer et al. 2013
							HDL	1.80E-04	185508	0.0188	0.0048	Willer et al. 2013
	rs769449	45410002	A	G	intron	APOE	LDL	4.90E-161	83660	0.2141	0.0076	Willer et al. 2013
							TC	6.80E-103	93260	0.1589	0.0072	Willer et al. 2013

Table S22. The 36 SNP pairs that were identified as having an additive effect on LDL.

Shown are the IDs of the SNPs in the Exome Chip (SNP1 and SNP2), the genes mapped close to the SNPs (Gene1 and Gene2), the beta-coefficient (Beta) and the p-value for the additive model. The SNP pairs that were hit in my co-RNAi screen are highlighted. (Data analysis performed by Gulum Kosova, unpublished)

	SNP1	Gene1	SNP2	Gene2	Beta	P-value
1	exm62588	PCSK9	exm.rs12740374	CELSR2	0.13	4.62E-07
2	exm62588	PCSK9	exm176096	APOB	0.11	5.19E-06
3	exm62588	PCSK9	exm.rs3846662	HMGCR	0.09	1.83E-05
4	exm62588	PCSK9	exm686341	LPL	0.17	4.90E-07
5	exm62588	PCSK9	exm.rs6511720	LDLR	0.17	8.64E-08
6	exm62588	PCSK9	exm.rs17216525	NCAN	0.18	3.60E-06
7	exm62588	PCSK9	exm1479366	APOE	0.44	2.12E-32
8	exm62588	PCSK9	exm.rs4420638	APOE	0.16	1.47E-08
9	exm.rs12740374	CELSR2	exm176096	APOB	0.1	1.04E-08
10	exm.rs12740374	CELSR2	exm.rs3846662	HMGCR	0.09	5.73E-08
11	**exm.rs12740374**	**CELSR2**	**exm686341**	**LPL**	**0.12**	**4.86E-09**
12	exm.rs12740374	CELSR2	exm.rs6511720	LDLR	0.13	8.95E-10
13	exm.rs12740374	CELSR2	exm.rs17216525	NCAN	0.12	4.22E-08
14	exm.rs12740374	CELSR2	exm1479366	APOE	0.22	1.12E-22
15	exm.rs12740374	CELSR2	exm.rs4420638	APOE	0.13	1.14E-10
16	**exm176096**	**APOB**	**exm.rs3846662**	**HMGCR**	**0.09**	**1.22E-07**
17	exm176096	APOB	exm686341	LPL	0.11	6.33E-08
18	exm176096	APOB	exm.rs6511720	LDLR	0.11	1.58E-08
19	exm176096	APOB	exm.rs17216525	NCAN	0.11	3.95E-07
20	exm176096	APOB	exm1479366	APOE	0.18	4.30E-19
21	exm176096	APOB	exm.rs4420638	APOE	0.11	2.30E-09
22	exm.rs3846662	HMGCR	exm686341	LPL	0.09	2.84E-07
23	exm.rs3846662	HMGCR	exm.rs6511720	LDLR	0.1	6.26E-08
24	exm.rs3846662	HMGCR	exm.rs17216525	NCAN	0.09	1.66E-06
25	exm.rs3846662	HMGCR	exm1479366	APOE	0.16	4.92E-17
26	exm.rs3846662	HMGCR	exm.rs4420638	APOE	0.1	8.16E-09
27	exm686341	LPL	exm.rs6511720	LDLR	0.14	2.61E-09
28	exm686341	LPL	exm.rs17216525	NCAN	0.14	7.85E-08
29	exm686341	LPL	exm.rs6511720	LDLR	0.14	2.61E-09
30	exm686341	LPL	exm.rs4420638	APOE	0.14	3.84E-10
31	exm.rs6511720	LDLR	exm.rs17216525	NCAN	0.15	1.57E-08
32	exm.rs6511720	LDLR	exm1479366	APOE	0.27	1.04E-25
33	exm.rs6511720	LDLR	exm.rs4420638	APOE	0.14	5.55E-11
34	exm.rs17216525	NCAN	exm1479366	APOE	0.3	2.75E-25
35	exm.rs17216525	NCAN	exm.rs4420638	APOE	0.15	1.56E-09
36	exm1479366	APOE	exm.rs4420638	APOE	0.22	2.16E-23

9.2 SUPPLEMENTARY FIGURES

Figure S1. Comparison of the knockdown effect on LDL uptake between the reverse transfection (screen) and the liquid-phase transfection

Shown is an example of the effect of gene knockdowns on LDL uptake, using the same siRNA amounts as in the screen (no siRNA adjustments) or smaller amounts of siRNAs than in the screen (adjusted siRNA concentration) (A), as well as an example of titration of an siRNA to determine the suitable amount to be used, in order to achieve the desirable phenotype.

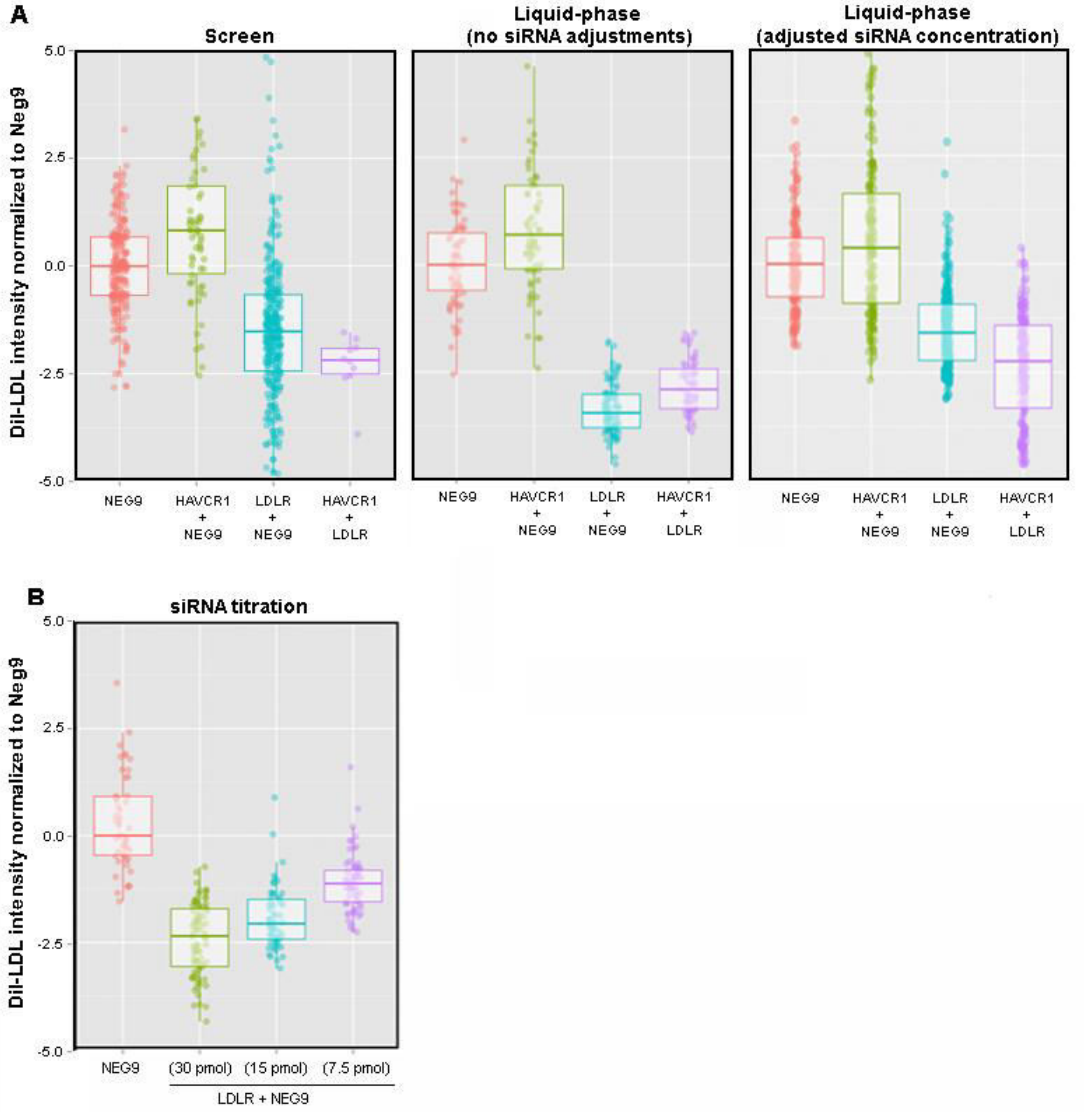
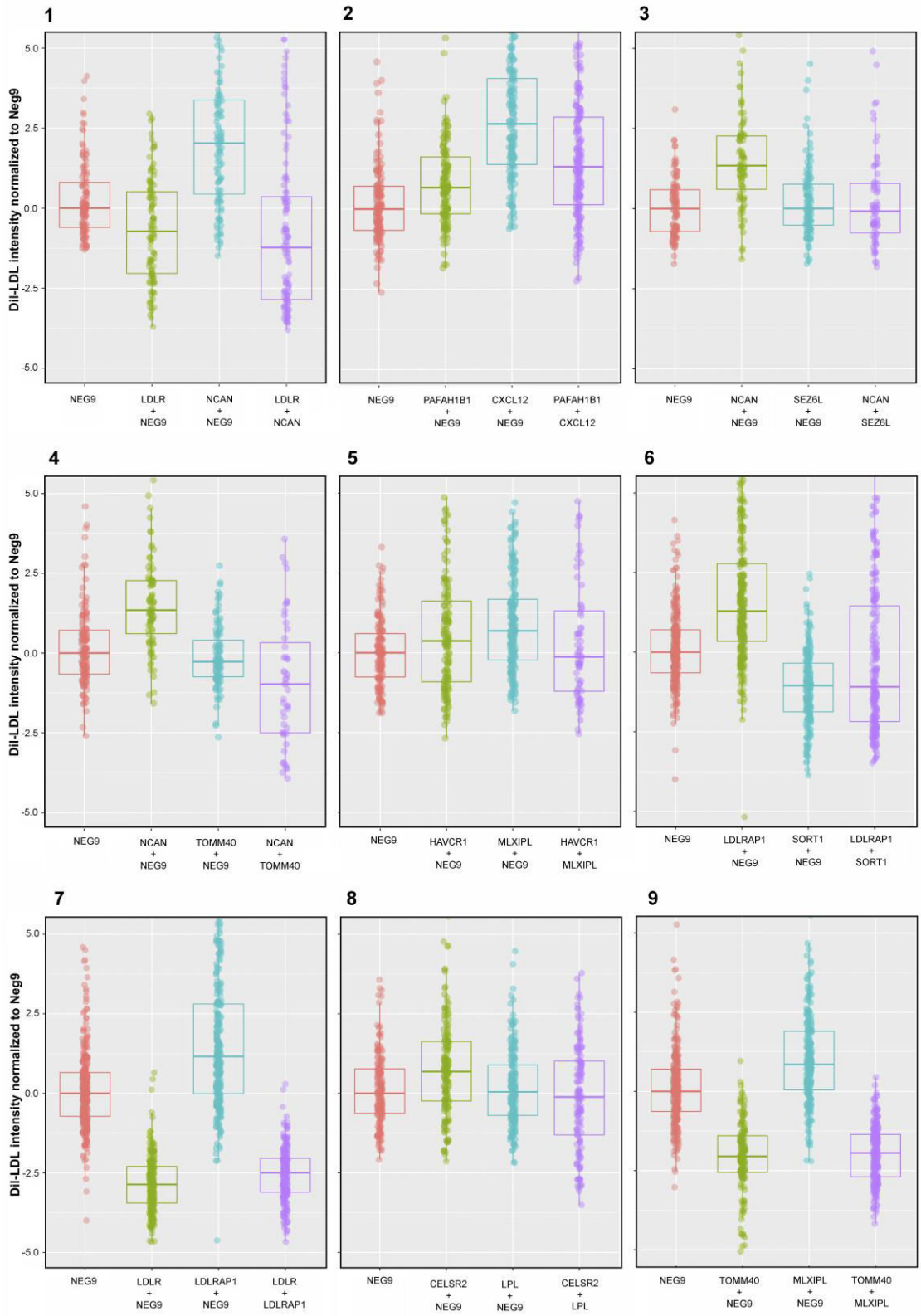
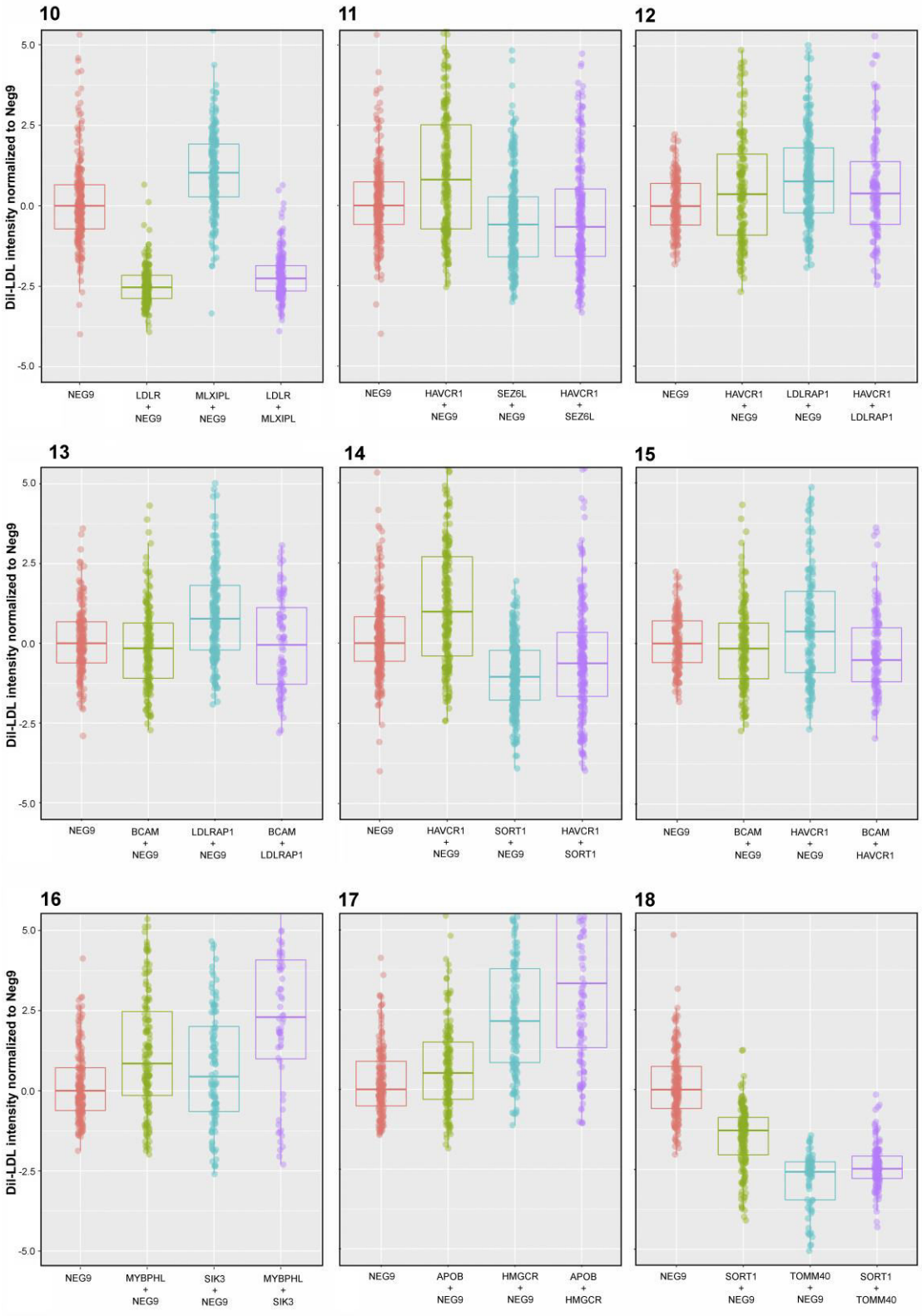
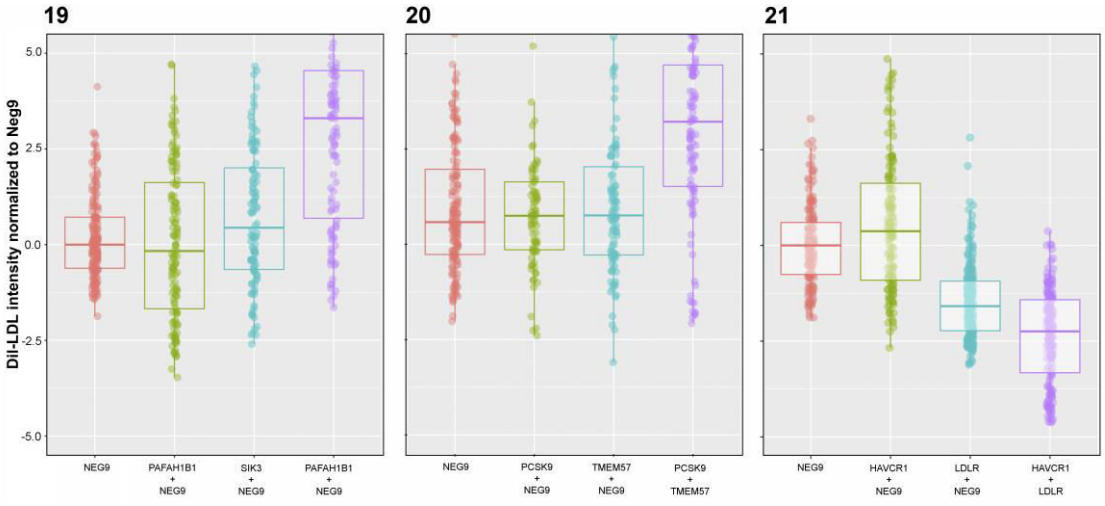


Figure S2. Boxplot representation of the 21 gene-gene interactions that were replicated with liquid-phase transfection.

Shown are the 21 gene-gene interactions that were validated with liquid-phase transfection. The boxplots show the median intensity of internalized DiI-LDL, normalized to the control, for the two single knockdowns (transfected together with the control siRNA), as well as for the double knockdown, for each pair of genes. (n=3-4)




10
11
12
13
14
15
16
17
18
Dil-LDL intensity normalized to Neg9
5.0
2.5
0.0
-2.5
-5.0
NEG9
LDLR + NEG9
MLXIPL + NEG9
LDLR + MLXIPL
HAVCR1 + NEG9
SEZ6L + NEG9
HAVCR1 + SEZ6L
LDLRAP1 + NEG9
HAVCR1 + LDLRAP1
BCAM + NEG9
LDLRAP1 + NEG9
BCAM + LDLRAP1
SORT1 + NEG9
HAVCR1 + SORT1
BCAM + HAVCR1
MYBPHL + NEG9
SIK3 + NEG9
MYBPHL + SIK3
APOB + NEG9
HMGCR + NEG9
APOB + HMGCR
SORT1 + NEG9
TOMM40 + NEG9
SORT1 + TOMM40


19
20
21
Dil-LDL intensity normalized to Neg9
5.0
2.5
0.0
-2.5
-5.0
NEG9
PAFAH1B1 + NEG9
SIK3 + NEG9
PAFAH1B1 + NEG9
NEG9
PCSK9 + NEG9
TMEM57 + NEG9
PCSK9 + TMEM57
NEG9
HAVCR1 + NEG9
LDLR + NEG9
HAVCR1 + LDLR

10 REFERENCES

Abifadel, M. et al. 2003. “Mutations in PCSK9 Cause Autosomal Dominant Hypercholesterolemia.” *Nature Genetics* 34(2):154–56.

Albert, F. W. and L. Kruglyak. 2015. “The Role of Regulatory Variation in Complex Traits and Disease.” *Nature Reviews Genetics* 16(4):197–212.

Altshuler, D., M. J. Daly, and E. Lander. 2009. “Genetic Mapping in Human Disease.” *Science* 322(5903):881–88.

Axelsson, E. et al. 2011. “Extracting Quantitative Genetic Interaction Phenotypes from Matrix Combinatorial RNAi.” *BMC Bioinformatics* 12(1):342.

Badano, J. L. and N. Katsanis. 2002. “Beyond Mendel: An Evolving View of Human Genetic Disease Transmission.” *Nature Reviews. Genetics* 3(10):779–89.

Barabasi, A. L. and Z. N. Oltvai. 2004. “Network Biology: Understanding the Cell’s Functional Organization.” *Nature Review* 5(February):101–13.

Barr, A. R. and C. Bakal. 2015. “A Sensitised RNAi Screen Reveals a Ch-TOG Genetic Interaction Network Required for Spindle Assembly.” *Scientific Reports* 5(April):10564.

Bartz, F. et al. 2009. “Identification of Cholesterol-Regulating Genes by Targeted RNAi Screening.” *Cell Metabolism* 10(1):63–75.

Baryshnikova, A., M. Costanzo, C. L. Myers, B. Andrews, and C. Boone. 2013. “Genetic Interaction Networks: Toward an Understanding of Heritability.” *Annual Review of Genomics and Human Genetics* 14(June):111–33.

Bassik, M. C. et al. 2013. “A Systematic Mammalian Genetic Interaction Map Reveals Pathways Underlying Ricin Susceptibility.” *Cell* 152(4):909–22.

Beltrao, P., G. Cagney, and N. J. Krogan. 2010. “Quantitative Genetic Interactions Reveal Biological Modularity.” *Cell* 141(5):739–45.

Benjamini, Y. and Y. Hochberg. 1995. “Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing.” *Source Journal of the Royal Statistical Society. Series B (Methodological) Journal of the Royal Statistical Society. Series B J. R. Statist. Soc. B* 57(1):289–300.

Billmann, M. et al. 2016. “A Genetic Interaction Map of Cell Cycle Regulators.” *Molecular Biology of the Cell* 27(8):1397–1407.

Birmingham, A. et al. 2009. “Statistical Methods for Analysis of High-Throughput RNA Interference Screens.” *Nature Methods* 6(8):569–75.

Blanchette-Mackie, E.Joan. 2000. “Intracellular Cholesterol Trafficking: Role of the NPC1 Protein.” *Biochimica et Biophysica Acta - Molecular and Cell Biology of Lipids* 1486(1):171–83.

Blattmann, P., C. Schuberth, R. Pepperkok, and H. Runz. 2013. “RNAi-Based Functional Profiling of Loci from Blood Lipid Genome-Wide Association Studies Identifies Genes with Cholesterol-Regulatory Function.” *PLoS Genetics* 9(2).

Boone, C., H. Bussey, and B. J. Andrews. 2007. “Exploring Genetic Interactions and Networks with Yeast. TL - 8.” *Nature Reviews. Genetics* 8 VN-re(6):437–49.

Botstein, D. and N. Risch. 2003. “Discovering Genotypes Underlying Human Phenotypes: Past Successes for Mendelian Disease, Future Approaches for Complex Disease.” *Nature Genetics* 33 Suppl(march):228–37.

Botstein, D., R. L. White, M. Skolnick, and R. W. Davis. 1980. “Construction of a Genetic Linkage Map in Man Using Restriction Fragment Length Polymorphisms.” *American Journal of Human Genetics* 32(3):314–31.

Brown, M. and JL Goldstein. 1986. “A Receptor-Mediated Pathway for Cholesterol Homeostasis.” *Science* 232(4746):34–47.

Brown, M. S. and J. L. Goldstein. 1974. “Familial Hypercholesterolemia: Defective Binding of Lipoproteins to Cultured Fibroblasts Associated with Impaired Regulation of 3-Hydroxy-3-Methylglutaryl Coenzyme A Reductase Activity.” *Proceedings of the National Academy of Sciences of the United States of America* 71(3):788–92.

Brown, M. S. and J. L. Goldstein. 1980. “Multivalent Feedback Regulation of HMG CoA Reductase, a Control Mechanism Coordinating Isoprenoid Synthesis and Cell Growth.” *Journal of Lipid Research* 21(44).

Brown, M. S. and J. L. Goldstein. 1997. “The SREBP Pathway: Regulation of Cholesterol Metabolism by

Proteolysis of a Membrane-Bound Transcription Factor." *Cell* 89(3):331–40.
Brown, M. S., J. Herz, and J. L. Goldstein. 1997. "LDL-Receptor Structure: Calcium Cages, Acid Baths and Recycling Receptors." *Nature*.
Brown, Michael S. and Joseph L. Goldstein. 2009. "Cholesterol Feedback: From Schoenheimer's Bottle to Scap's MELADL." *Journal of Lipid Research* 50 Suppl:S15–27.
Carlborg, O. and C. S. Haley. 2004. "Epistasis: Too Often Neglected in Complex Trait Studies?" *Nature Reviews. Genetics* 5(8):618–25.
Cha, J. Y. and J. J. Repa. 2007. "The Liver X Receptor (LXR) and Hepatic Lipogenesis: The Carbohydrate-Response Element-Binding Protein Is a Target Gene of LXR." *Journal of Biological Chemistry* 282(1):743–51.
Chetrit, D., N. Ziv, and M. Ehrlich. 2009. "Dab2 Regulates Clathrin Assembly and Cell Spreading." *The Biochemical Journal* 418(3):701–15.
Cirulli, E. T. E. T. and D. B. D. B. Goldstein. 2010. "Uncovering the Roles of Rare Variants in Common Disease through Whole-Genome Sequencing." *Nature Reviews. Genetics* 11(6):415–25.
Cohen, J. et al. 2005. "Low LDL Cholesterol in Individuals of African Descent Resulting from Frequent Nonsense Mutations in PCSK9." *Nature Genetics* 37(2):161–65.
Cohen, J. C., M. Kimmel, A. Polanski, and H. H. Hobbs. 2003. "Molecular Mechanisms of Autosomal Recessive Hypercholesterolemia." *Curr Opinion Lipidol* 14:121–27.
Cordell, H. J. 2002. "Epistasis: What It Means, What It Doesn't Mean, and Statistical Methods to Detect It in Humans." *Human Molecular Genetics* 11(20):2463–68.
Cordell, H. J. 2009. "Europe PMC Funders Group Detecting Gene-Gene Interactions That Underlie Human Diseases." *Nature Reviews. Genetics* 10(6):392–404.
Corella, D. et al. 2002. "Associations of LPL and APOC3 Gene Polymorphisms on Plasma Lipids in a Mediterranean Population: Interaction with Tobacco Smoking and the APOE Locus." *Journal of Lipid Research* 43(3):416–27.
Costanzo, M. et al. 2010. "The Genetic Landscape of a Cell." *Science (New York, N.Y.)* 327(5964):425–31.
Costanzo, M., A. Baryshnikova, C. L. Myers, B. Andrews, and C. Boone. 2011. "Charting the Genetic Interaction Map of a Cell." *Current Opinion in Biotechnology* 22(1):66–74.
Crow, J. F. 1990. "How Important Is Detecting Interaction?" *Behavioral and Brain Sciences* (13):126–127.
Cusick, M. E., N. Klitgord, M. Vidal, and D. E. Hill. 2005. "Interactome: Gateway into Systems Biology." *Human Molecular Genetics* 14(SUPPL. 2):171–81.
D'Agostino, R. B. et al. 2008. "General Cardiovascular Risk Profile for Use in Primary Care: The Framingham Heart Study." *Circulation* 117(6):743–53.
Dietschy, J. M., S. D. Turley, and D. K. Spady. 1993. "Role of Liver in the Maintenance of Cholesterol and Low Density Lipoprotein Homeostasis in Different Animal Species, Including Humans." *Journal of Lipid Research* 34(10):1637–59.
Dinu, I. et al. 2012. "SNP-SNP Interactions Discovered by Logic Regression Explain Crohn's Disease Genetics." *PLoS ONE* 7(10).
Do, R., S. Kathiresan, and G. R. Abecasis. 2012. "Exome Sequencing and Complex Disease: Practical Aspects of Rare Variant Association Studies." *Human Molecular Genetics* 21(R1):1–9.
Dobzhansky, T. 1965. "Bichromosomal Synthetic Semilethals in Drosophila Pseudoobscura." *Proceedings of the National Academy of Sciences of the United States of America* 482–86.
Eden, ER., X. M. Sun, Dilipkumar D. Patel, and Anne K. Soutar. 2007. "Adaptor Protein Disabled-2 Modulates Low Density Lipoprotein Receptor Synthesis in Fibroblasts from Patients with Autosomal Recessive Hypercholesterolaemia." *Human Molecular Genetics* 16(22):2751–59.
Eichler, E. 2010. "Missing Heritability and Strategies for Finding the Underlying Causes of Complex Disease." *Nat Rev Genet* 11(6):446–50.
Erfle, H. et al. 2007. "Reverse Transfection on Cell Arrays for High Content Screening Microscopy." *Nature Protocols* 2(2):392–99.
Erfle, H. et al. 2008. "Work Flow for Multiplexing siRNA Assays by Solid-Phase Reverse Transfection in Multiwell Plates." *Journal of Biomolecular Screening* 13(7):575–80.
Farhan, S. M. K. and R. A. Hegele. 2013. "Genetics 101 for Cardiologists: Rare Genetic Variants and Monogenic Cardiovascular Disease." *Canadian Journal of Cardiology* 29(1):18–22.
Favari, E. et al. 2015. "Cholesterol Efflux and Reverse Cholesterol Transport." Pp. 181–206 in *Handbook of Experimental Pharmacology*, vol. 224.

Fisher, A. R. 1918. “The Correlation between Relatives on the Supposition of Mendelian Inheritance.” *Transactions of the Royal Society of Edinburgh* (December):273–83.
Flannick, Jason and Jose C. Florez. 2016. “Type 2 Diabetes: Genetic Data Sharing to Advance Complex Disease Research.” *Nature Reviews Genetics* 17(9):535–49.
Garcia, C. K. 2001. “Autosomal Recessive Hypercholesterolemia Caused by Mutations in a Putative LDL Receptor Adaptor Protein.” *Science* 292(5520):1394–98.
Garuti, R. et al. 2005. “The Modular Adaptor Protein Autosomal Recessive Hypercholesterolemia (ARH) Promotes Low Density Lipoprotein Receptor Clustering into Clathrin-Coated Pits.” *Journal of Biological Chemistry* 280(49):40996–4.
Geiger, T., A. Wehner, C. Schaab, J. Cox, and M. Mann. 2012. “Comparative Proteomic Analysis of Eleven Common Cell Lines Reveals Ubiquitous but Varying Expression of Most Proteins.” *Molecular & Cellular Proteomics* 11(3):M111.014050-M111.014050.
Gelissen, I. C. and A. J. McLachlan. 2014. “The Pharmacogenomics of Statins.” *Pharmacological Research* 88:99–106.
Giaever, G. et al. 2002. “Functional Profiling of the Saccharomyces Cerevisiae Genome.” *Nature* 418(6896):387–91.
Gilbert-Diamond, D. and J. H. Moore. 2011. “Analysis of Gene-Gene Interactions.” *Curr Protoc Hum Genet. 2011* (5):724–32.
Gilbert, D. F., T. Meinhof, R. Pepperkok, and H. Runz. 2009. “DetecTiff(C): A Novel Image Analysis Routine for High-Content Screening Microscopy.” *J Biomol Screen* 1087057109339523.
Glass, C. K. and J. L. Witztum. 2001. “Atherosclerosis : The Road Ahead Review.” *Cell* 104(4):503–16.
Goldstein, D. B. et al. 2013. “Sequencing Studies in Human Genetics: Design and Interpretation.” *Nature Reviews. Genetics* 14(7):460–70.
Goldstein, J. L. and M. S. Brown. 1977. “THE LOW-DENSITY Lipoprotein Pathway and Its RELATION TO ATHEROSCLEROSIS.” *Atherosclerosis* 897–930.
Goldstein, J. L. and M. S. Brown. 2015. “A Century of Cholesterol and Coronaries: From Plaques to Genes to Statins.” *Cell* 161(1):161–72.
Goldstein, J. L., M. S. Brown, R. G. W. Anderson, D. W. Russell, and W. J. Schneider. 1985. “RECEPTOR-MEDIATED ENDOCYTOSIS: Concepts Emerging from the LDL Receptor System.” *Receptor* 1–39.
Goldstein, J. L., RA. DeBose-Boyd, and M. S. Brown. 2006. “Protein Sensors for Membrane Sterols.” *Cell* 124(1):35–36.
Gould, A. L., J. E. Rossouw, N. C. Santanello, J. F. Heyse, and C. D. Furberg. 1998. “Impact of Statin Trials.” *Circulation* 97:946–52.
Grundberg, E. et al. 2012. “Mapping Cis- and Trans-Regulatory Effects across Multiple Tissues in Twins.” *Nature Genetics* 44(10):1084–89.
Guo, X., D. C. Trudgian, A. Lemoff, S. Yadavalli, and H. Mirzaei. 2014. “Confetti: A Multiprotease Map of the HeLa Proteome for Comprehensive Proteomics.” *Molecular & Cellular Proteomics : MCP* 13(6):1573–84.
Haines, J. et al. 2005. “Complement Factor H Variant Increases the Risk of Age-Related Macular Degeneration.” *Science (New York, NY)* 308(5720):419.
Hall, M. A., J. H. Moore, and M. D. Ritchie. 2016. “Embracing Complex Associations in Common Traits: Critical Considerations for Precision Medicine.” *Trends in Genetics* 32(8):470–84.
Harvey, R. and D. Ferrier. 2011. *Lippincott’s Illustrated Reviews: Biochemistry*. 5th ed. Lippincott Wlliams and Wilkins.
He, G. et al. 2002. “ARH Is a Modular Adaptor Protein That Interacts with the LDL Receptor, Clathrin, and AP-2.” *Journal of Biological Chemistry* 277(46):44044–49.
Horn, T. et al. 2011. “Mapping of Signaling Networks through Synthetic Genetic Interaction Analysis by RNAi.” *Nature Methods* 8(4):341–46.
Hosseini, H. et al. 2015. “Phosphatidylserine Liposomes Mimic Apoptotic Cells to Attenuate Atherosclerosis by Expanding Polyreactive IgM Producing B1a Lymphocytes.” *Cardiovascular Research* 106(3):443–52.
Ichimura, T. et al. 2008. “Kidney Injury Molecule-1 Is a Phosphatidylserine Receptor That Confers a Phagocytic Phenotype on Epithelial Cells.” *Journal of Clinical Investigation* 118(5):1657–68.
Ikonen, E. 2008. “Cellular Cholesterol Trafficking and Compartmentalization.” *Nature Reviews. Molecular Cell Biology* 9(2):125–38.
International Human Genome Sequencing Consortium. 2004. “Finishing the Euchromatic Sequence of the Human Genome.” *Nature* 50(2):162–68.

Kampmann, Martin, Michael C. Bassik, and Jonathan S. Weissman. 2013. “Integrated Platform for Genome-Wide Screening and Construction of High-Density Genetic Interaction Maps in Mammalian Cells. TL - 110.” *Proceedings of the National Academy of Sciences of the United States of America* 110 VN-(25):26.

Kardia, S. L. et al. 2006. “Epistatic Effects between Two Genes in the Renin-Angiotensin System and Systolic Blood Pressure and Coronary Artery Calcification.” *Med Sci Monit* 12(4):CR150-8.

Kathiresan, S. et al. 2009. “UKPMC Funders Group Six New Loci Associated with Blood Low-Density Lipoprotein Cholesterol , High-Density Lipoprotein Cholesterol or Triglycerides in Humans.” 40(2):189–97.

Kathiresan, S. and D. Srivastava. 2012. “Genetics of Human Cardiovascular Disease.” *Cell* 148(6):1242–57.

Kathiresan, S., CJ Willer, GM Peloso, and S. Demissie. 2008. “Common Variants at 30 Loci Contribute to Polygenic Dyslipidemia.” *Nat Genet* 41(1):56–65.

Keys, A. 1966. “Epidemiological Studies Related To Coronary Heart Disease: Characteristics of Men Aged 40-59 in Seven Countries*.” *Acta Med Scand Suppl*.

Khera, Amit V and Sekar Kathiresan. 2017. “Genetics of Coronary Artery Disease: Discovery, Biology and Clinical Translation.” *Nature Publishing Group*.

Kiezun, A. et al. 2013. “Exome Sequencing and the Genetic Basis of Complex Traits.” *NIH Public Access* 44(6):623–30.

Krainova, N. A. et al. 2013. “Evaluation of Potential Reference Genes for qRT-PCR Data Normalization in HeLa Cells.” *Applied Biochemistry and Microbiology* 49(9):743–49.

Kruglyak, L. and D. a Nickerson. 2001. “Variation Is the Spice of Life.” *Nature Genetics* 27(3):234–36.

Kuroda, M. et al. 2015. “Interaction between TIM-1 and NPC1 Is Important for Cellular Entry of Ebola Virus.” *Journal of Virology* 89(12):6481–93.

Lagor, WR and JS Millar. 2010. “Overview of the LDL Receptor: Relevance to Cholesterol Metabolism and Future Approaches for the Treatment of Coronary Heart Disease.” *J Receptor Ligand Channel Res*.

Landry, J. J. M. et al. 2013. “The Genomic and Transcriptomic Landscape of a HeLa Cell Line.” *G3 (Bethesda, Md.)* 3(8):1213–24.

Lanktree, M. B. and R. a Hegele. 2009. “Gene-Gene and Gene-Environment Interactions: New Insights into the Prevention, Detection and Management of Coronary Artery Disease.” *Genome Medicine* 1(2):28.

Laufer, C., B. Fischer, M. Billmann, W. Huber, and M. Boutros. 2013. “Mapping Genetic Interactions in Human Cancer Cells with RNAi and Multiparametric Phenotyping.” *Nature Methods* 10(5):427–31.

Laufer, C., B. Fischer, W. Huber, and M. Boutros. 2014. “Measuring Genetic Interactions in Human Cells by RNAi and Imaging.” *Nat Protoc* 9(10):2341–53.

Lehner, Ben, Julia Tischler, and Andrew G. Fraser. 2006. “RNAi Screens in Caenorhabditis Elegans in a 96-Well Liquid Format and Their Application to the Systematic Identification of Genetic Interactions.” *Nature Protocols* 1(3):1617–20.

Lehrman, M. 1985. “Mutation in LDL Receptor: Alu-Alu Recombination Deletes Exons Encoding Transmembrane and Cytoplasmic Domains.” 227(4):395–401.

Lepor, N. E. and D. J. Kereiakes. 2015. “The PCSK9 Inhibitors: A Novel Therapeutic Target Enters Clinical Practice.” *American Health & Drug Benefits* 8(9):483–89.

Lin, H. Y. et al. 2013. “SNP-SNP Interaction Network in Angiogenesis Genes Associated with Prostate Cancer Aggressiveness.” *PLoS ONE* 8(4).

Linsel-Nitschke, P., N. J. Samani, and H. Schunkert. 2010. “Sorting out Cholesterol and Coronary Artery Disease.” *The New England Journal of Medicine* 363(25):2462–63.

Lucas, G. et al. 2012. “Hypothesis-Based Analysis of Gene-Gene Interactions and Risk of Myocardial Infarction.” *PLoS ONE* 7(8).

Lusis, A. J. 2012. “Genetics of Atherosclerosis.” *Trends in Genetics* 28(6):267–75.

Maher, Brendan. 2008. “The Case of the Missing Heritability.” *Nature* 456(November):18–21.

Malo, N., J. A. Hanley, S. Cerquozzi, J. Pelletier, and R. Nadon. 2006. “Statistical Practice in High-Throughput Screening Data Analysis.” *Nat Biotechnol* 24(2):167–75.

Mani, R., RP St Onge, J. L. Hartman, G. Giaever, and F. P. Roth. 2008. “Defining Genetic Interaction.” *Proceedings of the National Academy of Sciences of the United States of America* 105(9):3461–66.

Manolio, TA. 2009. “Finding the Missing Heritability of Complex Diseases.” *Nature* 461(7265):747–53.
Markowetz, F. and M. Boutros. 2015. *Systems Genetics*. edited by F. Markowetz and M. Boutros. Cambridge University Press.
Maurer, M. E. and J. A. Cooper. 2006. “The Adaptor Protein Dab2 Sorts LDL Receptors into Coated Pits Independently of AP-2 and ARH.” *Journal of Cell Science* 119(Pt 20):4235–46.
Mayeux, R. 2005. “Review Series Introduction Mapping the New Frontier : Complex Genetic Disorders.” *The Journal of Clinical Investigation* 115(6).
McCarthy, M. I. and JN. Hirschhorn. 2008. “Genome-Wide Association Studies: Potential next Steps on a Genetic Journey.” *Human Molecular Genetics* 17(R2):156–65.
McPherson, R. and A. Tybjaerg-Hansen. 2016. “Genetics of Coronary Artery Disease.” *Circulation Research* 118(4):564–78.
Mishra, S. K. et al. 2002. “Disabled-2 Exhibits the Properties of a Cargo-Selective Endocytic Clathrin Adaptor.” *EMBO Journal* 21(18):4915–26.
Mortensen, M. B. et al. 2014. “Targeting Sortilin in Immune Cells Reduces Proinflammatory Cytokines and Atherosclerosis.” *Journal of Clinical Investigation* 124(12):5317–22.
Muntendam, P., C. McCall, J. Sanz, E. Falk, and V. Fuster. 2010. “The BioImage Study: Novel Approaches to Risk Assessment in the Primary Prevention of Atherosclerotic Cardiovascular Disease-Study Design and Objectives.” *American Heart Journal* 160(1):49–57.e1.
Musameh, M. D. et al. 2015. “Analysis of Gene-Gene Interactions among Common Variants in Candidate Cardiovascular Genes in Coronary Artery Disease.” *PLoS ONE* 10(2):1–12.
Musunuru, Kiran et al. 2011. “From Noncoding Variant to Phenotype via SORT1 at the 1p13 Cholesterol Locus.” *Nature* 466(7307):714–19.
Musunuru, Kiran. 2015. “Personalized Genomes and Cardiovascular Disease.” *Cold Spring Harbor Perspectives in Medicine* 5(1):1–10.
Natarajan, P. et al. 2015. “Association of APOC3 Loss-of-Function Mutations with Plasma Lipids and Subclinical Atherosclerosis: The Multi-Ethnic BioImage Study.” *Journal of the American College of Cardiology* 66(18):2053–55.
Nelson, J. K. et al. 2015. “Deubiquitylase USP2 Regulates the LDLR Pathway by Counteracting the E3-Ubiquitin Ligase IDOL.” *Circulation Research* 118(3):410–19.
Neumann, B. et al. 2006. “High-Throughput RNAi Screening by Time-Lapse Imaging of Live Human Cells.” *Nature Methods* 3(5):385–90.
Neumann, B. et al. 2011. “Phenotypic Profiling of the Human Genome by Time-Lapse Microscopy Reveals Cell Division Genes.” *Nature* 464(7289):721–27.
Nielsen, M. S., C. Jacobsen, G. Olivecrona, C. M. Petersen, and J. Gliemann. 1999. “Sortilin/neurotensin Receptor-3 Binds Lipoprotein Lipase and Mediates Its Degradation.” *Atherosclerosis* 144(13):8832–36.
Nijman, S. M. B. 2011. “Synthetic Lethality: General Principles, Utility and Detection Using Genetic Screens in Human Cells.” *FEBS Letters* 585(1):1–6.
Onay, V. U. et al. 2006. “SNP-SNP Interactions in Breast Cancer Susceptibility.” *BMC Cancer* 6(1):114.
Park, M. J. et al. 2014. “High Glucose-Induced O-GlcNAcylated Carbohydrate Response Element-Binding Protein (ChREBP) Mediates Mesangial Cell Lipogenesis and Fibrosis the Possible Role in the Development of Diabetic Nephropathy.” *Journal of Biological Chemistry* 289(19):13519–30.
Patel, K. M. et al. 2015. “Macrophage Sortilin Promotes LDL Uptake, Foam Cell Formation, and Atherosclerosis.” *Circulation Research* 116(5):789–96.
Patel, S. B. et al. 1998. “Mapping a Gene Involved in Regulating Dietary Cholesterol Absorption.” *Journal of Clinical Investigation* 1041–44.
Peloso, G. M. et al. 2014. “Association of Low-Frequency and Rare Coding-Sequence Variants with Blood Lipids and Coronary Heart Disease in 56,000 Whites and Blacks.” *American Journal of Human Genetics* 94(2):223–32.
Pepperkok, R. and J. Ellenberg. 2006. “High-Throughput Fluorescence Microscopy for Systems Biology.” *Nature Reviews Molecular Cell Biology* 7(9):690–696.
Pérez-Pérez, JM, H. Candela, and J. L. Micol. 2009. “Understanding Synergy in Genetic Interactions.” *Trends in Genetics* 25(8):368–76.
Perron, P. et al. 2007. “Apolipoprotein E and Lipoprotein Lipase Gene Polymorphisms Interaction on the Atherogenic Combined Expression of Hypertriglyceridemia and Hyperapobetalipoproteinemia Phenotypes.” *Journal of Endocrinological Investigation* 30(7):551–57.

Phillips, P. C. 1998. “The Language of Gene Interaction.” *Genetics* 149(3):1167–71.
Phillips, PC. 2008. “Epistasis—the Essential Role of Gene Interactions in the Structure and Evolution of Genetic Systems.” *Nature Reviews Genetics* 9(11):855–67.
Pirucello, J. and S. Kathiresan. 2010. “Genetics of Lipid Disorders.” *Current Opinion in Cardiology* 25(3):238–42.
Pitas, R. E., T. L. Innerarity, J. N. Weinstein, and R. W. Mahley. 1981. “Acetoacetylated Lipoproteins Used to Distinguish Fibroblasts from Macrophages in Vitro by Fluorescence Microscopy.” *Arteriosclerosis (Dallas, Tex.)* 1(3):177–85.
Pritchard, J. K. 2001. “Are Rare Variants Responsible for Susceptibility to Complex Diseases?” *American Journal of Human Genetics* 69(1):124–37.
Puliti, A., G. Caridi, R. Ravazzolo, and G. M. Ghiggeri. 2007. “Teaching Molecular Genetics: Chapter 4 - Positional Cloning of Genetic Disorders.” *Pediatric Nephrology* 22(12):2023–29.
Rauch, U., K. Feng, and X. H. Zhou. 2001. “Neurocan: A Brain Chondroitin Sulfate Proteoglycan.” *Cellular and Molecular Life Sciences : CMLS* 58(12–13):1842–56.
Reich, D. E. and E. S. Lander. 2001. “On the Allelic Spectrum of Human Disease.” *Trends in Genetics* 17(9):502–10.
Ribeiro, I. et al. 2001. “Niemann-Pick Type C Disease: NPC1 Mutations Associated with Severe and Mild Cellular Cholesterol Trafficking Alterations.” *Human Genetics* 109(1):24–32.
Ritzerfeld, J. et al. 2011. “Phenotypic Profiling of the Human Genome Reveals Gene Products Involved in Plasma Membrane Targeting of SRC Kinases Phenotypic Profiling of the Human Genome Reveals Gene Products Involved in Plasma Membrane Targeting of SRC Kinases.” 1955–68.
Roguev et al. 2013. “Quantitative Genetic-Interaction Mapping in Mammalian Cells.” *Nature Methods* 10(5):432–37.
Rolland, T. et al. 2014. “A Proteome-Scale Map of the Human Interactome Network.” *Cell* 159(5):1212–26.
Rual, J. F., K. Venkatesan, and T. Hao. 2005. “Towards a Proteome-Scale Map of the Human Protein – Protein Interaction Network.” *Nature* 437(October):1173–78.
Russell, D. W. 1992. “Cholesterol Biosynthesis and Metabolism.” *Cardiovascular Drugs and Therapy / Sponsored by the International Society of Cardiovascular Pharmacotherapy* 6(2):103–10.
Samani, N. J., J. Erdmann, A. S. Hall, and C. Hengstenberg. 2007. “Genomewide Association Analysis of Coronary Artery Disease.” *The New England Journal of Medicine* 443–53.
Sarrazy, V. et al. 2015. “Maintenance of Macrophage Redox Status by ChREBP Limits Inflammation and Apoptosis and Protects against Advanced Atherosclerotic Lesion Formation.” *Cell Reports* 13(1):132–44.
Sawa, A. and S. H. Snyder. 2002. “Schizophrenia: Diverse Approaches to a Complex Disease.” *Science* 296(5568):692–95.
Schork, N. J., S. S. Murray, K. A. Frazer, and E. J. Topol. 2010. “Common vs. Rare Allele Hypotheses for Complex Diseases.” 19(3):212–19.
Schuldiner, M. et al. 2005. “Exploration of the Function and Organization of the Yeast Early Secretory Pathway through an Epistatic Miniarray Profile.” *Cell* 123(3):507–19.
Shimano, H. et al. 1997. “Isoform 1c of Sterol Regulatory Element Binding Protein Is Less Active than Isoform 1a in Livers of Transgenic Mice and in Cultured Cells.” *Journal of Clinical Investigation* 99(5):846–54.
Simpson, J. C. et al. 2012. “Genome-Wide RNAi Screening Identifies Human Proteins with a Regulatory Function in the Early Secretory Pathway.” *Nature Cell Biology* 14(7):764–74.
Singleton, Andrew and John Hardy. 2016. “The Evolution of Genetics: Alzheimer’s and Parkinson’s Diseases.” *Neuron* 90(6):1154–63.
Sirinian, MI et al. 2005. “Adaptor Protein ARH Is Recruited to the Plasma Membrane by Low Density Lipoprotein (LDL) Binding and Modulates Endocytosis of the LDL/LDL Receptor Complex in Hepatocytes.” *Journal of Biological Chemistry* 280(46):38416–23.
Smith, D. J. and A. J. Lusis. 2002. “The Allelic Structure of Common Disease.” *Human Molecular Genetics* 11(20):2455–61.
Song, B. L., N. B. Javitt, and R. A. DeBose-Boyd. 2005. “Insig-Mediated Degradation of HMG CoA Reductase Stimulated by Lanosterol, an Intermediate in the Synthesis of Cholesterol.” *Cell Metabolism* 1(3):179–89.
Soria, L. F. et al. 1989. “Association between a Specific Apolipoprotein B Mutation and Familial Defective Apolipoprotein B-100 (Genetic Disease/cholesterol Metabolism).” *Genetics*

86(January):587–91.
Spassky, B., T. Dobzhansky, and W. W. Anderson. 1965. “Genetics of Natural Populations. XXXVI. Epistatic Interactions of the Components of the Genetic Load in Drosophila Pseudoobscura.” *Genetics* 52(3):653–64.
Stoeckman, A. K., L. Ma, and H. C. Towle. 2004. “Mlx Is the Functional Heteromeric Partner of the Carbohydrate Response Element-Binding Protein in Glucose Regulation of Lipogenic Enzyme Genes.” *Journal of Biological Chemistry* 279(15):15662–69.
Strong, A. et al. 2012. “Hepatic Sortilin Regulates Both Apolipoprotein B Secretion and LDL Catabolism.” *Journal of Clinical Investigation* 122(8):2807–16.
Tada, H. et al. 2014. “Multiple Associated Variants Increase the Heritability Explained for Plasma Lipids and Coronary Artery Disease.” *Circulation: Cardiovascular Genetics* 7(5):583–87.
Tall, A. R. and D. Ai. 2011. “Sorting out Sortilin.” *Circulation Research* 108(2):158–60.
Teslovich, T., K. Musunuru, and A. Smith. 2010. “Biological, Clinical and Population Relevance of 95 Loci for Blood Lipids.” *Nature* 466(7307):707–13.
The 1000 Genomes Project Consortium. 2015. “A Global Reference for Human Genetic Variation.” *Nature* 526(7571):68–74.
The International HapMap Consortium. 2003. “The International HapMap Project.” *Nature* 426(6968):789–96.
The International HapMap Consortium. 2005. “A Haplotype Map of the Human Genome.” *Nature* 437(7063):1299–320 ST–A haplotype map of the human genome.
The International HapMap Consortium. 2007. “A Second Generation Human Haplotype Map of over 3.1 Million SNPs.” *Nature* 449(7164):851–61.
Tietjen, Gregory T. et al. 2014. “Molecular Mechanism for Differential Recognition of Membrane Phosphatidylserine by the Immune Regulatory Receptor Tim4.” *Proceedings of the National Academy of Sciences of the United States of America* 111(15):E1463-72.
Tischler, J., B. Lehner, N. Chen, and A. G. Fraser. 2006. “Combinatorial RNA Interference in Caenorhabditis Elegans Reveals That Redundancy between Gene Duplicates Can Be Maintained for More than 80 Million Years of Evolution.” *Genome Biology* 7(8):R69.
Tong, AH Y. 2004. “Global Mapping of the Yeast Genetic Interaction Network.” *Science* 303(October):808–13.
Tsai, CT et al. 2007. “Renin-Angiotensin System Gene Polymorphisms and Coronary Artery Disease in a Large Angiographic Cohort: Detection of High Order Gene-Gene Interaction.” *Atherosclerosis* 195(1):172–80.
Wang, W., Bryan J. B. J. Barratt, David G. D. G. Clayton, and J. A.John A. Todd. 2005. “Genome-Wide Association Studies: Theoretical and Practical Concerns.” *Nature Reviews Genetics* 6(2):109–18.
Wang, Xiaoyue, Audrey Q. Fu, Megan E. McNerney, and Kevin P. White. 2014. “Widespread Genetic Epistasis among Cancer Genes.” *Nature Communications* 5:4828.
Wei, W. H., G. Hemani, and C. S. Haley. 2014. “Detecting Epistasis in Human Complex Traits.” *Nat Rev Genet* 15(11):722–33.
Willer, C. J. et al. 2013. “Discovery and Refinement of Loci Associated with Lipid Levels.” *Nature Genetics* 45(11):1274–83.
Zeggini, E. and A. Morris. 2010. *Analysis of Complex Disease Association Studies*. Academic Press Inc.
Zelcer, N. 2009. “LXR Regulates Cholesterol Uptake through Idol-Dependent Ubiquitination of the LDL Receptor.” *Science* 141(4):520–29.
Zhang, D. W. et al. 2007. “Binding of Proprotein Convertase Subtilisin/kexin Type 9 to Epidermal Growth Factor-like Repeat A of Low Density Lipoprotein Receptor Decreases Receptor Recycling and Increases Degradation.” *Journal of Biological Chemistry* 282(25):18602–12.
Zhang, L. et al. 2011. “The IDOL-UBE2D Complex Mediates Sterol-Dependent Degradation of the LDL Receptor.” *Genes and Development* 25(12):1262–74.
Zhang, X., J. A. Kuivenhoven, and A. K. Groen. 2015. “Forward Individualized Medicine from Personal Genomes to Interactomes.” *Frontiers in Physiology* 6(DEC):1–11.
Zuk, O. et al. 2014. “Searching for Missing Heritability: Designing Rare Variant Association Studies.” *Proceedings of the National Academy of Sciences of the United States of America* 111(4):E455-64.
Zuk, O., E. Hechter, SR Sunyaev, and ES Lander. 2012. “The Mystery of Missing Heritability: Genetic Interactions Create Phantom Heritability.” *Proceedings of the National Academy of Sciences of the United States of America* 109(4):1193–98.