Circulation Research

American
Heart
Association®

Mutations in *STAP1* Are Associated With Autosomal Dominant Hypercholesterolemia

Sigrid W. Fouchier, Geesje M. Dallinga-Thie, Joost C.M. Meijers, Noam Zelcer, John J.P. Kastelein, Joep C. Defesche and G. Kees Hovingh

The online version of this article, along with updated information and services, is located on the
World Wide Web at:
http://circres.ahajournals.org/content/115/6/552

Data Supplement (unedited) at:
http://circres.ahajournals.org/content/suppl/2014/07/16/CIRCRESAHA.115.304660.DC1.html

Mutations in *STAP1* Are Associated With Autosomal Dominant Hypercholesterolemia

Sigrid W. Fouchier, Geesje M. Dallinga-Thie, Joost C.M. Meijers, Noam Zelcer,
John J.P. Kastelein, Joep C. Defesche, G. Kees Hovingh

Rationale: Autosomal-dominant hypercholesterolemia (ADH) is characterized by elevated low-density lipoprotein cholesterol levels and increased risk for coronary vascular disease. ADH is caused by mutations in the low-density lipoprotein receptor, apolipoprotein B, or proprotein convertase subtilisin/kexin 9. A number of patients, however, suffer from familial hypercholesterolemia 4 (FH4), defined as ADH in absence of mutations in these genes and thereafter use the abbreviation FH4.

Objective: To identify a fourth locus associated with ADH.

Methods and Results: Parametric linkage analysis combined with exome sequencing in a FH4 family resulted in the identification of the variant p.Glu97Asp in signal transducing adaptor family member 1 (*STAP1*), encoding signal transducing adaptor family member 1. Sanger sequencing of *STAP1* in 400 additional unrelated FH4 probands identified a second p.Glu97Asp carrier and 3 additional missense variants, p.Leu69Ser, p.Ile71Thr, and p.Asp207Asn. *STAP1* carriers (n=40) showed significantly higher plasma total cholesterol and low-density lipoprotein cholesterol levels compared with nonaffected relatives (n=91).

Conclusions: We mapped a novel ADH locus at 4p13 and identified 4 variants in *STAP1* that associate with ADH.   (*Circ Res*. 2014;115:552-555.)

Key Words: hypercholesterolemia, autosomal dominant ■ physical chromosome mapping ■ *STAP1* gene, human

Autosomal-dominant hypercholesterolemia (ADH) is characterized by elevated low-density lipoprotein cholesterol (LDL-C) levels and increased risk for coronary vascular disease. ADH is a common disease with a prevalence of ≈1 in 200 individuals in most Western countries.[1] To date, mutations in 3 genes, LDL receptor (*LDLR*), specific domains of apolipoprotein B (*APOB*), and proprotein convertase subtilisin/kexin 9 (*PCSK9*), have been shown to cause ADH.[2–5] Although mutations in these genes can be found in a large proportion of clinically diagnosed ADH patients, the underlying molecular determinant remains unknown in a substantial fraction of patients.[3] Unraveling these molecular determinants is a major focus of cardiovascular research, yet identification of novel ADH genes is largely challenged by the phenotypic and genetic heterogeneity of this disease.[6,7]

We describe *STAP1*, encoding signal transducing adaptor family member 1, as a fourth gene associated with ADH.

Editorial, see p 534
In This Issue, see p 533

Methods

A detailed description of the experimental procedure and statistical analysis is provided in the Online Data Supplement.

Briefly, we identified a family affected with ADH in which we first excluded the presence of causal *LDLR*, *APOB*, and *PCSK9* mutations in the proband (Figure; familial hypercholesterolemia 4 [FH4]-0102-III-12). We then obtained genetic mapping data to identify shared haplotypes, phenocopies and nonpenetrants, and selected the most distantly related and informative individuals for exome sequencing. Exome sequencing was performed for 3 affected family members (FH4-0102-III-12, FH4-0102-IV-4, and FH4-0102-IV-11).

Results

Parametric linkage analysis revealed the highest possible logarithm of odds scores when individual FH4-0102-III-7, FH4-0102-III-8, and FH4-0102-III-11 were considered to be phenocopies, resulting in highly suggestive LOD scores of 3.0 (θ=0.0) at chromosomes 4p15.1-q13.3, 11p13-q14.1, and 13q14.13-q32.1 (Online Figure I). Exome sequencing was performed for 3 affected family members with shared haplotypes at the 3 identified linkage intervals. Several discrete filtering

steps (Online Table I) of the exome sequence data led to the identification of a variant in *STAP1*, p.Glu97Asp (c.291G>C). This variant, which was verified by Sanger sequencing, affected a highly conserved amino acid and was predicted to be pathogenic by SIFT, Polyphen-2, and MutationTaster. The variant was absent in 400 controls of matched ancestry, the Genome of the Netherlands (http://www.nlgenome.nl/), the 1000G population, and the National Heart, Lung, and Blood Institute-Exome Sequencing Project data set (http://evs.gs.washington.edu/EVS/), suggesting that it is not a polymorphism. Testing of remaining family members by Sanger sequencing revealed 11 affected carriers (total cholesterol or LDL-C levels >95th percentile), 4 possibly affected carriers (between 75th and 95th

percentile) and 1 unaffected carrier (<75th percentile), reflecting a disease penetrance of 0.94 (Figure).

We subsequently sequenced all coding regions of *STAP1* in 400 unrelated FH4 probands (Online Table II) and identified a second p.Glu97Asp carrier and 3 additional missense variants, p.Leu69Ser (c.206T>C), p.Ile71Thr (c.212T>C), and p.Asp207Asn (c.619G>A). All variants were located in exon 3, except for p.Asp207Asn located in exon 6, and all variants altered highly conserved amino acid residues. SIFT, Polyphen-2, and MutationTaster predicted the p.Leu69Ser variant to be pathogenic, whereas p.Ile71Thr and p.Asp207Asn were predicted to possibly affect protein function. None of these variants were detected in 400 controls, the 1000G population, or in the Genome of the Netherlands. In the National Heart, Lung, and Blood Institute-Exome Sequencing Project data set, p.Ile71Thr was found once and p.Asp207Asn was found 3×. However, variants at low frequency associated with an hypercholesterolemic phenotype can be expected, considering the substantial number of patients at increased cardiovascular risk included in National Heart, Lung, and Blood Institute-Exome Sequencing Project.

The frequency of rare *STAP1* variants predicted to be pathogenic or possibly pathogenic in our FH4 cohort was 1.3% (5 of 400 individuals). This was 3× and 6× higher than the rate observed in National Heart, Lung, and Blood Institute-Exome Sequencing Project (0.4%, 17 of 4300 European American
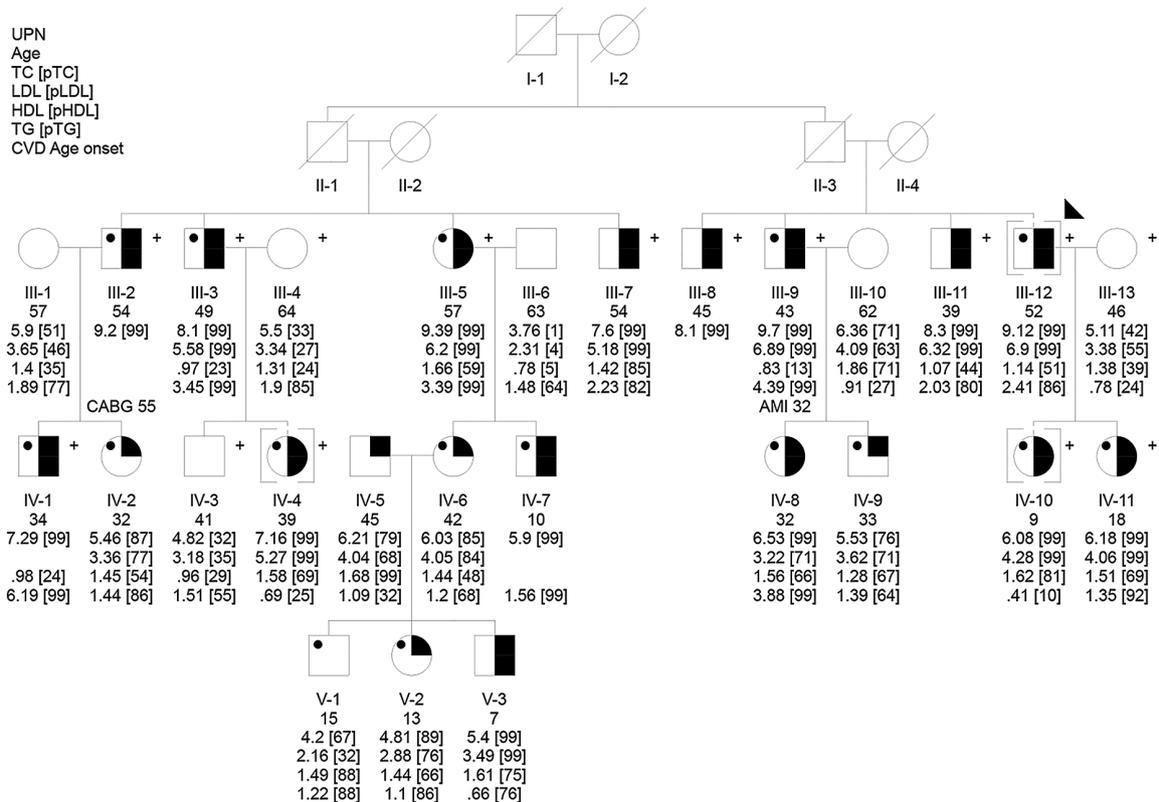
Figure. Family FH4-0102. Half-blackened symbols indicate affected members (>95th percentile), quarter-blackened symbols indicate possibly affected members (between 75th and 95th percentile), and white symbols indicate unaffected members (<75th percentile). A dot indicates carriers of the p.Glu97Asp variant. The triangle represents the proband. Age (in years) at lipid measurement, total cholesterol (TC), low-density lipoprotein cholesterol (LDL-C), high-density lipoprotein cholesterol (HDL-C), triglycerides (TG) in mmol/L, and history and age onset of coronary vascular disease are given. Percentiles matched for age and sex, pTC, pLDL, pHDL, and pTG are given between brackets, a percentile indicated by 99 resembles a >95th percentile. Individuals with mapping data are marked with +. For individuals with exome data symbols are marked with brackets. AMI indicates acute myocardial infarction; CABG, coronary artery bypass graft; CVD, coronary vascular disease; and UPN, unique personal number.

individuals; *P*=0.035) and the Genome of the Netherlands (0.2%, 1 of 499 individuals; *P*=0.067), respectively (Online Table III).

The *STAP1* variants found in the 4 additional probands were further evaluated in the relatives of the replicate families. In total, 45 carriers of the rare *STAP1* variants and 91 noncarrier relatives were identified. Individuals carrying a rare *STAP1* variant had significantly higher total cholesterol and LDL-C levels compared with relatives without the *STAP1* variants (Table). High-density lipoprotein cholesterol levels were not significantly different, whereas triglycerides levels were slightly increased. Similar results were observed when the carriers of the discovery family were excluded (Online Table IV).

We additionally compared the ADH phenotype of the *STAP1* carriers with the ADH phenotype observed in a large cohort of *LDLR* and *APOB* mutation carriers. Total cholesterol and LDL-C levels were significantly lower in *STAP1* carriers when compared with patients molecularly diagnosed with familial hypercholesterolemia (FH, OMIM 606945), caused by mutations in *LDLR*. However, cholesterol levels in *STAP1* carriers were comparable with those observed in patients molecularly diagnosed with familial defective apolipoproteinemia B (OMIM 107730), caused by mutations in *APOB*.

In studies looking at ADH, individuals are considered affected if their plasma cholesterol levels are >95th percentile adjusted for age and sex, and in ADH-affected families a disease penetrance of 0.9 is usually assumed. Nevertheless, the ADH phenotype of patients affected by well-established pathogenic mutations in the *LDLR* and *APOB* varies enormously from severe to moderate, as recently demonstrated.[8,9] In the 5 *STAP1* families identified here, such variation was also observed, most evident in family FH4-0150 (Online Figure IIA), which displayed only a moderate penetrance of 0.76. The variants p.Ile71Thr in family FH4-0247 (Online Figure IIB) and p.Glu97Asp in family FH4-317 (Online Figure IIC) displayed complete penetrance. Further expansion of family FH4-0356 was unfortunately not possible (Online Figure IID). Overall, the disease penetrance of the missense *STAP1* variants in all 5 families is 0.94±0.10 with a phenocopy rate of 0.19±0.11.

Discussion

We report here for the first time association of rare *STAP1* variants with increased total cholesterol and LDL-C levels. The lipid phenotype of *STAP1* carriers is not as severe as generally observed in FH patients, but resembles the minor lipid and coronary vascular disease phenotype observed in familial defective apolipoproteinemia B patients.[4] It is of note that in genome-wide association studies, single nucleotide polymorphism in the STAP1 locus have not been shown to be associated with cholesterol levels. However, single nucleotide polymorphisms in the STAP1 locus have been associated with the risk for Parkinson disease.[10] Although the exact biological explanation for this association remains elusive, increased cholesterol levels might play a role because high total serum cholesterol levels have been associated with a slow clinical progression of Parkinson disease.[11]

The function of STAP1, also known as BRDG1 (BCR downstream signaling protein 1) or stem cell adaptor protein 1, is largely unknown. STAP1 contains a Pleckstrin homology domain, a Src homology 2 domain, and several tyrosine phosphorylation sites.[12] It has been suggested that the Pleckstrin homology domain of STAP1 functions as a phosphoinositide-binding domain and facilitates the association of STAP1 with membranes.[12] Intriguingly, the p.Leu69Ser, p.Ile71Thr, and p.Glu97Asp alterations all occur within the Pleckstrin homology domain of STAP1 and may thus affect its interaction with membranes or membrane proteins (Online Figure III). The *STAP1* p.Asp207Asn variant is located in the Src homology 2 domain, a domain implicated in propagation of signal transduction emanating from upstream receptor tyrosine kinase pathways.[13] Indeed, STAP1 has been suggested to act downstream of the receptor tyrosine kinases *c-kit* and *c-fms*.[12] In line with STAP1 functioning downstream of *c-kit* to control systemic cholesterol levels, it is interesting to point out that W/W[v] mice, which harbor a loss-of-function mutation in *c-kit*, show increased plasma cholesterol levels.[14,15] Similarly, gain-of-function mutations in *c-kit* cause leukemia,[16] and leukemic patients often present with hypocholesterolemia.[17–21] Intriguingly, therapeutic administration of receptor tyrosine kinase inhibitors to these patients is associated with increases in plasma cholesterol.[22,23] This raises the possibility that *STAP1* and potentially other downstream genes in this signaling pathway have an effect on circulating levels of cholesterol.

Our study provides strong genetic evidence for involvement of STAP1 in controlling cholesterol homeostasis, yet the molecular mechanism behind this is presently unknown. Expression of *STAP1* in the liver, the main organ maintaining whole-body cholesterol homeostasis, is low (not shown). Studies aimed at identifying the extrahepatic mechanism underlying the potential

Table. Clinical Characteristics of Carriers and Noncarriers

	n	STAP1 Noncarriers,* mean [SD]	n	STAP1 Carriers, mean [SD]	P Value	n	APOB Carriers, mean [SD]	STAP1 vs APOB, P Value	n	LDLR Carriers, mean [SD]	STAP1 vs LDLR, P Value
Sex (men), %	91	43	40	50	...	2471	49	...	13234	48	...
Age, y	91	48 [16]	40	38 [19]	...	2471	40 [21]	...	13234	38 [21]	...
TC, mmol/L	80	5.76 [1.25]	40	6.63 [1.66]	0.001	2471	7.17 [2.13]	0.432	13234	8.44 [2.80]	<0.001
LDL, mmol/L	89	3.83 [1.10]	35	4.27 [1.24]	0.005	1963	4.76 [1.57]	0.075	9176	5.37 [1.98]	<0.001
HDL, mmol/L	78	1.37 [0.37]	36	1.33 [0.28]	0.949	1970	1.18 [0.36]	0.009	9247	1.17 [0.36]	0.001
TG, mmol/L	78	1.41 [0.79]	37	1.85 [1.32]	0.052	1969	1.30 [0.85]	0.006	9201	1.30 [0.85]	0.011

APOB indicates apolipoprotein B; HDL, high-density lipoprotein cholesterol; LDL, low-density lipoprotein cholesterol; *LDLR*, low-density lipoprotein receptor; *STAP1*, signal transducing adaptor family member 1; TC, total cholesterol; and TG, triglycerides.

*All family members, including branches in which the *STAP1* variants were not present.

role of STAP1 in cholesterol metabolism are required and are expected to yield new insight into the pathogenesis of ADH.

References

1. Nordestgaard BG, Chapman MJ, Humphries SE, et al; European Atherosclerosis Society Consensus Panel. Familial hypercholesterolaemia is underdiagnosed and undertreated in the general population: guidance for clinicians to prevent coronary heart disease: consensus statement of the European Atherosclerosis Society. *Eur Heart J*. 2013;34:3478–390a.
2. Fouchier SW, Defesche JC, Umans-Eckenhausen MW, Kastelein JP. The molecular basis of familial hypercholesterolemia in The Netherlands. *Hum Genet*. 2001;109:602–615.
3. Fouchier SW, Kastelein JJ, Defesche JC. Update of the molecular basis of familial hypercholesterolemia in The Netherlands. *Hum Mutat*. 2005;26:550–556.
4. Fouchier SW, Defesche JC, Kastelein JJ, Sijbrands EJ. Familial defective apolipoprotein B versus familial hypercholesterolemia: an assessment of risk. *Semin Vasc Med*. 2004;4:259–264.
5. Abifadel M, Varret M, Rabès JP, et al. Mutations in PCSK9 cause autosomal dominant hypercholesterolemia. *Nat Genet*. 2003;34:154–156.
6. Varret M, Abifadel M, Rabès JP, Boileau C. Genetic heterogeneity of autosomal dominant hypercholesterolemia. *Clin Genet*. 2008;73:1–13.
7. Altmüller J, Palmer LJ, Fischer G, Scherb H, Wjst M. Genomewide scans of complex human diseases: true linkage is hard to find. *Am J Hum Genet*. 2001;69:936–950.
8. Huijgen R, Kindt I, Fouchier SW, Defesche JC, Hutten BA, Kastelein JJ, Vissers MN. Functionality of sequence variants in the genes coding for the low-density lipoprotein receptor and apolipoprotein B in individuals with inherited hypercholesterolemia. *Hum Mutat*. 2010;31:752–760.
9. Huijgen R, Hutten BA, Kindt I, Vissers MN, Kastelein JJ. Discriminative ability of LDL-cholesterol to identify patients with familial hypercholesterolemia: a cross-sectional study in 26,406 individuals tested for genetic FH. *Circ Cardiovasc Genet*. 2012;5:354–359.
10. Fung HC, Scholz S, Matarin M, et al. Genome-wide genotyping in Parkinson's disease and neurologically normal controls: first stage analysis and public release of data. *Lancet Neurol*. 2006;5:911–916.
11. Huang X, Auinger P, Eberly S, Oakes D, Schwarzschild M, Ascherio A, Mailman R, Chen H; Parkinson Study Group DATATOP Investigators. Serum cholesterol and the progression of Parkinson's disease: results from DATATOP. *PLoS One*. 2011;6:e22854.
12. Masuhara M, Nagao K, Nishikawa M, Sasaki M, Yoshimura A, Osawa M. Molecular cloning of murine STAP-1, the stem-cell-specific adaptor protein containing PH and SH2 domains. *Biochem Biophys Res Commun*. 2000;268:697–703.
13. Koytiger G, Kaushansky A, Gordus A, Rush J, Sorger PK, MacBeath G. Phosphotyrosine signaling proteins that drive oncogenesis tend to be highly interconnected. *Mol Cell Proteomics*. 2013;12:1204–1213.
14. Hatanaka K, Tanishita H, Ishibashi-Ueda H, Yamamoto A. Hyperlipidemia in mast cell-deficient W/WV mice. *Biochim Biophys Acta*. 1986;878:440–445.
15. Smith ER, Yeasky T, Wei JQ, Miki RA, Cai KQ, Smedberg JL, Yang WL, Xu XX. White spotting variant mouse as an experimental model for ovarian aging and menopausal biology. *Menopause*. 2012;19:588–596.
16. Blume-Jensen P, Hunter T. Oncogenic kinase signalling. *Nature*. 2001;411:355–365.
17. Gilbert HS, Ginsberg H, Fagerstrom R, Brown WV. Characterization of hypocholesterolemia in myeloproliferative disease. Relation to disease manifestations and activity. *Am J Med*. 1981;71:595–602.
18. Naik PP, Ghadge MS, Raste AS. Lipid profile in leukemia and Hodgkin's disease. *Indian J Clin Biochem*. 2006;21:100–102.
19. Yavasoglu I, Tombuloglu M, Kadikoylu G, Donmez A, Cagirgan S, Bolaman Z. Cholesterol levels in patients with multiple myeloma. *Ann Hematol*. 2008;87:223–228.
20. Allampallam K, Dutt D, Nair C, Shetty V, Mundle S, Lisak L, Andrews C, Ahmed B, Mazzone L, Zorat F, Borok R, Muzammil M, Gundroo A, Ansaarie I, Raza A. The clinical and biologic significance of abnormal lipid profiles in patients with myelodysplastic syndromes. *J Hematother Stem Cell Res*. 2000;9:247–255.
21. Pandolfino J, Hakimian D, Rademaker AW, Tallman MS. Hypocholesterolemia in hairy cell leukemia: a marker for proliferative activity. *Am J Hematol*. 1997;55:129–133.
22. Kim TD, Rea D, Schwarz M, et al. Peripheral artery occlusive disease in chronic phase chronic myeloid leukemia patients treated with nilotinib or imatinib. *Leukemia*. 2013;27:1316–1321.
23. Rea D, Mirault T, Cluzeau T, Gautier JF, Guilhot F, Dombret H, Messas E. Early onset hypercholesterolemia induced by the 2nd-generation tyrosine kinase inhibitor nilotinib in patients with chronic phase-chronic myeloid leukemia. *Haematologica*. 2014;99:1197–1203.

Novelty and Significance

What Is Known?

- Autosomal-dominant hypercholesterolemia (ADH) leads to an increase in total and low-density lipoprotein cholesterol and accelerated atherosclerotic disease.
- Mutations in the genes encoding low-density lipoprotein receptor, apolipoprotein B, or proprotein convertase subtilisin/kexin 9 are established causes of ADH.
- A substantial proportion of clinically diagnosed hypercholesterolemic patients do not carry mutations in the 3 established causal genes for ADH.
- The identification of novel ADH genes is essential to identifying additional targets to lower low-density lipoprotein cholesterol to reduce coronary vascular disease risk.

What New Information Does This Article Contribute?

- We identified signal transducing adaptor family member 1 (*STAP1*) as a novel gene associated with ADH.

- This finding reveals a previously unrecognized signaling pathway that contributes to the pathogenesis of ADH.
- Identification of *STAP1* as an ADH gene might expand current genetic strategies for the molecular diagnosis of ADH patients and lead to new therapeutic targets for ADH.

The prevalence of ADH is ≈1:200 individuals in most Western countries. The diagnosis is usually made on the basis of clinical symptoms. However, because these symptoms develop late in life, establishing the diagnosis in younger patients is often difficult. Genetic analysis, that is, the demonstration of a causative mutation, provides unequivocal diagnosis. A proportion of the patients with a clinical diagnosis of ADH does not carry mutations in the established ADH genes but might carry mutations in genes that have not been associated previously with the hypercholesterolemic phenotype. Here, we identify *STAP1* as a novel ADH gene, a finding of prognostic value that could also yield new insight into the pathogenesis of ADH.

SUPPLEMENTAL MATERIAL

DETAILED METHODS

Family recruitment

The probands were clinically diagnosed with familial hypercholesterolemia using an uniform protocol and internationally accepted criteria[1,2] and were screened for the presence of mutations in *LDLR, APOB* (amino acids 2955 to 3092, 3303 to 3492, 3441 to 3615 and 4330 to 4440), and *PCSK9* as described before.[3,4]

Relatives were included into the study after written informed consent. The study protocol was approved by AMC's institutional review board (MEC 07/331).

After an overnight fast, blood was sampled and plasma levels of total cholesterol, HDL-cholesterol and triglycerides were measured by commercially available kits (Diasys, Holzheim, Germany and Randox, Northern Ireland, United Kingdom). LDL-c levels were calculated by the Friedewald formula only when triglycerides were below 4.5 mmol/L.[5]

An affected ADH phenotype was defined by levels of TC or LDL-c above the 95th percentile for age and gender.[6] A possible ADH phenotype was defined by levels of TC or LDL-c between the 75th and 95th percentile for age and gender.[7,8] When untreated values were not available, estimated baseline LDL-c values were back calculated based on the potency of cholesterol-lowering therapy.[9]

Linkage analysis

Genomic DNA was extracted from 10 ml whole blood on an AutopureLS apparatus according to a protocol provided by the manufacturer (Gentra Systems, Minneapolis, MI, USA).

Fifteen individuals from Family FH4-0102 (III:2, III:3, III:4, III:5, III:7, III:8, III:9, III:11, III:12, III:13, IV:1, IV:3, IV:4, IV:10, IV:11) were genotyped using the HumanCytoSNP-12 SNP array (Illumina). To verify the relationship between individuals, the data were subjected to the PedCheck program.[10] Multi-point linkage analysis was performed using SNPs with a spacing of 0.1 cM, computed in overlapping sets of 175 and 200 markers. Multi-point linkage analysis was first performed based on a model defined by a disease allele frequency of 0.001 and an autosomal dominant mode of inheritance with the assumption of 90% penetrance and 10% phenocopy rate using Genehunter[11] and Allegro[12]. The Allegro haplotypes were further studied in detail using HaploPainter[13] to identify phenocopies, non-penetrants and shared haplotypes.

The high penetrance of affected individuals in the 3nd generation of family FH4-0102 was unexpected and discordant with the expected Mendelian inheritance pattern. This suggested that two different ADH traits might be present in this family. Therefore, the three phenocopies identified in this generation were excluded and parametric LOD scores were re-calculated based on a model defined by a disease allele frequency of 0.001 and an autosomal dominant mode of inheritance with the assumption of 100% penetrance and 0% phenocopy using Genehunter.[11]

Exome sequencing

Whole exome sequencing was performed at the Broad Institute (Boston, USA). Briefly, genomic DNA underwent library construction and in-solution hybrid selection as previously described[14] using an Agilent (Agilent Technologies, Santa Clara, CA, USA) whole exome kit targeting 33Mb of genomic sequence. The target includes 188,260 exons from 18,560 genes. The resulting exomic DNA was sequenced using 75-base pair paired-end reads on an Illumina GA-II sequencer.[14] Raw sequence reads were aligned to the human genome reference (HG19) using the Burroughs-Wheeler Alignment tool[15] in paired-end mode. The aligned reads underwent base quality recalibration using the Genome Analysis Toolkit (GATK).[16,17] Duplicated reads from sequencing the same DNA molecule and sequencing reads not corresponding to the exome target were discarded.

Single nucleotide substitutions and short insertion and deletion events were identified and genotyped using the GATK UnifiedGenotyper tool in multisample mode. The quality scores of the resulting variants were recalibrated using the GATK Variant Score Recalibration.

Cohorts

By the participation of 64 Lipid Clinics, evenly distributed across The Netherlands, a representative group of over ~17,000 clinically diagnosed ADH patients has been collected. All participants gave written informed consent. From this cohort 400 unrelated cases, without any mutation in *LDLR, APOB* (amino acids 2955 to 3092, 3303 to 3492, 3441 to 3615 and 4330 to 4440) and *PCSK9* and with baseline TC and LDL-c levels above the 95th percentile for age and gender, were selected for further study.

The control cohort of ancestry-matched individuals and the cohorts of *APOB* and *LDLR* mutation carriers were recruited via the national genetic cascade screening program for FH.[18] The control cohort consisted of 400 unrelated and untreated individuals, with baseline LDL-c levels below the 20th percentile for age and gender and tested negative for the familial mutation. The cohorts of *APOB* and *LDLR* mutation carriers consisted of related and unrelated individuals tested positive for the familial mutation. Probands were excluded. When baseline values of the *APOB* and *LDLR* carriers were not available, estimated baseline LDL values were back calculated based on the potency of cholesterol-lowering therapy.[9] All subjects gave written informed consent for genetic screening. Lipid profiles were routinely measured with the LDX analyzer.[19]

The National Heart, Lung, and Blood Institute (NHLBI) Exome Sequencing Project (ESP) and the Genome of the Netherlands (GoNL)[20] datasets were used to evaluate the prevalence of individuals carrying rare *STAP1* variants.

Genetic analysis of *STAP1*

The nine exons, including the intronic boundaries, of *STAP1* were analyzed by Sanger sequencing. For mutation nomenclature, numbering was based on the cDNA with nucleotide c.1 being A of the ATG initiation codon p.1.[21] The *STAP1* reference sequence NM_000384.2 was used.

A variant predicted to be tolerated by SIFT[22], benign by Polyphen-2[23] and a polymorphism by MutationTaster[24] was considered to be non-pathogenic. A variant predicted to be deleterious by SIFT, Polyphen-2 and MutationTaster was considered to be pathogenic. All other variants were classified as possible pathogenic.

Phenocopy rate was determined by the count of non-carriers with an affected status divided by the count of all non-carriers in each family. Penetrance rate was determined by the count of carriers with an affected status divided by the count of all carriers in each family. Family members introduced by marriage were hereby excluded.

Statistical analysis

SPSS 20 software (SPSS, Inc, Chicago, Illinois, USA) was used for data analysis. Differences in cholesterol levels corrected for age, gender, and relatedness were determined using a general linear mixed model analysis. Triglyceride levels values were log-transformed before analysis. Probands were excluded from the analysis. A *p*-value <0.05 was considered statistically significant.

Differences in the frequency of rare *STAP1* variants between the Dutch FH4 cohort, GoNL and EVS-EA were determined using an one-tailed Fisher's exact test (GraphPad Software, La Jolla, California, USA)

References

1. Defesche JC. Familial Hypercholesterolemia. In: D.J.Betteridge, M.Dunitz, eds. *Lipids and Vascular Disease.* Ltd.London; 2000. p. 65-76.
2. Goldstein JL, Hobbs HH, Brown MS. Familial Hypercholesterolemia. In: Scriver CR, Beaudet AL, Sly WS, Valle D, eds. *The metabolic and molecular basis of inherited disease.* New York: McGraw-Hill; 1995. p. 1981-2030.

3. Fouchier SW, Kastelein JJ, Defesche JC. Update of the molecular basis of familial hypercholesterolemia in The Netherlands. *Hum Mutat*. 2005;26:550-6.
4. Motazacker MM, Pirruccello J, Huijgen R, Do R, Gabriel S, Peter J, Kuivenhoven JA, Defesche JC, Kastelein JJ, Hovingh GK, Zelcer N, Kathiresan S, Fouchier SW. Advances in genetics show the need for extending screening strategies for autosomal dominant hypercholesterolaemia. *Eur Heart J*. 2012.
5. Friedewald WT, Levy RI, Fredrickson DS. Estimation of the concentration of low-density lipoprotein cholesterol in plasma, without use of the preparative ultracentrifuge. *Clin Chem*. 1972;18:499-502.
6. Gotto AM, Jr., Bierman EL, Connor WE, Ford CH, Frantz ID, Jr., Glueck CJ, Grundy SM, Little JA. Recommendations for treatment of hyperlipidemia in adults. A joint statement of the Nutrition Committee and the Council on Arteriocslerosis. *Circulation*. 1984;69:1065A-90A.
7. Huijgen R, Kindt I, Fouchier SW, Defesche JC, Hutten BA, Kastelein JJ, Vissers MN. Functionality of sequence variants in the genes coding for the low-density lipoprotein receptor and apolipoprotein B in individuals with inherited hypercholesterolemia. *Hum Mutat*. 2010;31:752-60.
8. Huijgen R, Hutten BA, Kindt I, Vissers MN, Kastelein JJ. Discriminative ability of LDL-cholesterol to identify patients with familial hypercholesterolemia: a cross-sectional study in 26,406 individuals tested for genetic FH. *Circ Cardiovasc Genet*. 2012;5:354-9.
9. Walma EP, Visseren FL, Jukema JW, Kastelein JJ, Hoes AW, Stalenhoef AF. [The practice guideline 'Diagnosis and treatment of familial hypercholesterolaemia' of the Dutch Health Care Insurance Board]. *Ned Tijdschr Geneeskd*. 2006;150:18-23.
10. O'Connell JR, Weeks DE. PedCheck: a program for identification of genotype incompatibilities in linkage analysis. *Am J Hum Genet*. 1998;63:259-66.
11. Kruglyak L, Daly MJ, Reeve-Daly MP, Lander ES. Parametric and nonparametric linkage analysis: a unified multipoint approach. *Am J Hum Genet*. 1996;58:1347-63.
12. Gudbjartsson DF, Jonasson K, Frigge ML, Kong A. Allegro, a new computer program for multipoint linkage analysis. *Nat Genet*. 2000;25:12-3.
13. Thiele H, Nurnberg P. HaploPainter: a tool for drawing pedigrees with complex haplotypes. *Bioinformatics*. 2005;21:1730-2.
14. Fisher S, Barry A, Abreu J et al. A scalable, fully automated process for construction of sequence-ready human exome targeted capture libraries. *Genome Biol*. 2011;12:R1.
15. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2009;25:1754-60.
16. DePristo MA, Banks E, Poplin R et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet*. 2011;43:491-8.
17. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res*. 2010;20:1297-303.
18. Umans-Eckenhausen MA, Defesche JC, Sijbrands EJ, Scheerder RL, Kastelein JJ. Review of first 5 years of screening for familial hypercholesterolaemia in the Netherlands. *Lancet*. 2001;357:165-8.
19. Cobbaert C, Boerma GJ, Lindemans J. Evaluation of the Cholestech L.D.X. desktop analyser for cholesterol, HDL-cholesterol, and triacylglycerols in heparinized venous blood. *Eur J Clin Chem Clin Biochem*. 1994;32:391-4.
20. Boomsma DI, Wijmenga C, Slagboom EP et al. The Genome of the Netherlands: design, and project goals. *Eur J Hum Genet*. 2014;22:221-7.
21. den Dunnen JT, Antonarakis SE. Mutation nomenclature extensions and suggestions to describe complex mutations: a discussion. *Hum Mutat*. 2000;15:7-12.
22. Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc*. 2009;4:1073-81.
23. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, Kondrashov AS, Sunyaev SR. A method and server for predicting damaging missense mutations. *Nat Methods*. 2010;7:248-9.
24. Schwarz JM, Rodelsperger C, Schuelke M, Seelow D. MutationTaster evaluates disease-causing potential of sequence alterations. *Nat Methods*. 2010;7:575-6.

Supplementary note

Online Table I. Variant-filtering pipeline for family FH4-0102

criteria	# exome variants
Variants fitting a dominant model located in the linked intervals on chromosome 4, 11 and 13	462
Variants with frequency <0.5% (EVS and GoNL)	44
Variants not found in control exomes (GQ>90)	9
Not located in a gene with nonsense mutations with frequency >0.5% (stop-gained, frame-shift, and splice-3/5 in EVS and GoNL)	5
Not involving 1) non-coding variants predicted not to affect splicing and low phastcons (<0.5), 2) synonymous coding variants predicted not to affect splicing or 3) coding variants predicted to be non-pathogenic by SIFT, polyphen-2 and MutationTaster	1

Online Table II. Genetic variation in *STAP1* in 400 FH4 patients

exon	rs number	c.Nomen	p.Nomen	AA	AB	BB	F
1		c.12G>A	p.= (p.Lys4Lys)	387	1	0	0.001
1	rs187909999	c.120+6T>C	p.?	381	7	0	0.009
2	rs13111555	c.192+10T>C	p.?	9	71	297	0.882
3	rs2306810	c.193-12T>C	p.?	195	140	16	0.245
3		c.206T>C	p.Leu69Ser	336	1	0	0.001
3	rs141647940	c.212T>C	p.Ile71Thr	336	1	0	0.001
3		c.291G>C	p.Glu97Asp	336	2	0	0.003
4	rs41278563	c.307-167C>T	p.?	223	153	18	0.240
4	rs11556614	c.354A>G	p.= (p.Thr118Thr)	63	170	160	0.623
5	rs11556615	c.381C>T	p.= (p.Asn127Asn)	211	160	28	0.271
5	rs149803575	c.414G>C	p.= (p.Leu138Leu)	394	1	0	0.001
5	rs2242331	c.530+11A>G	p.?	186	176	36	0.312
5		c.530+13T>A	p.?	394	1	0	0.001
5	rs2242330	c.530+60A>G	p.?	244	135	19	0.217
6	rs202089275	c.531-16dup	p.?	327	45	1	0.063
6	rs146545610	c.619G>A	p.Asp207Asn	372	1	0	0.001
8	rs3816821	c.730-41G>C	p.?	92	177	103	0.515
8	rs192899556	c.826+66T>C	p.?	371	1	0	0.001
9		c.827-53C>T	p.?	392	1	0	0.001

Online Table III. Genetic variation in de coding region of *STAP1* indentified in GoNL and EVS-EA

Chr:Position	c.Nomen	p.Nomen	Function	FH4	GoNL	EVS-EA	S	P-2	MT	Causality
4:68424532	c.5T>A	p.Met2Lys	missense			1	D	D	D	pathogenic
4:68424562	c.35G>A	p.Arg12His	missense			1	D	D	D	pathogenic
4:68424573	c.46C>T	p.Gln16*	nonsense			1	N/A	N/A	N/A	pathogenic
4:68424619	c.92T>A	p.Phe31Tyr	missense			1	T	B	D	possible pathogenic
4:68441116	c.206T>C	p.Leu69Ser	missense	1			D	D	D	pathogenic
4:68441122	c.212T>C	p.Ile71Thr	missense	1		1	D	B	D	possible pathogenic
4:68441143	c.233C>T	p.Thr78Ile	missense			1	T	B	D	possible pathogenic
4:68441201	c.291G>C	p.Glu97Asp	missense	2			D	D	D	pathogenic
4:68447022	c.364-1G>A	p.?	splicing			1	N/A	N/A	N/A	pathogenic
4:68447185	c.526C>T	p.Pro176Ser	missense		1	3	D	D	D	pathogenic
4:68449354	c.596dup	p.Asn199Lysfs*8	frame-shift			1	N/A	N/A	N/A	pathogenic
4:68449380	c.619G>A	p.Asp207Asn	missense	1		3	T	B	D	possible pathogenic
4:68449386	c.625A>G	p.Arg209Gly	missense			1	T	B	P	non-pathogenic
4:68456651	c.709A>G	p.Thr237Ala	missense			1	D	D	D	pathogenic
4:68459021	c.773A>G	p.Lys258Arg	missense			1	T	D	D	possible pathogenic
4:68472019	c.832G>A	p.Glu278Lys	missense			1	T	D	P	possible pathogenic

S:SIFT; P-2:Polyphen-2; MT:MutationTaster

Online Table IV. Clinical characteristics of non-carriers versus carriers of the replicate families.

	N	STAP1 non-carriers* means [SD]	N	STAP1 carriers means [SD]	P
sex (% male)	91	43	25	50	
Age (years)	91	48 [16]	25	42 [19]	
TC (mmol/L)	80	5.76 [1.25]	25	6.54 [1.67]	0.003
LDL (mmol/L)	89	3.83 [1.10]	23	4.25 [1.17]	0.007
HDL (mmol/L)	78	1.37 [0.37]	23	1.31 [0.29]	0.587
TG (mmol/L)	78	1.41 [0.79]	23	1.60 [0.98]	0.370

TC:total cholesterol; LDL:LDL-cholesterol; HDL:HDL-cholesterol; TG:triglycerides;. *All family members, including branches in which the *STAP1* variants were not present.

Supplementary note

Online Figure I. Parametric LOD scores obtained in family FH4-0102

(A)

(B)
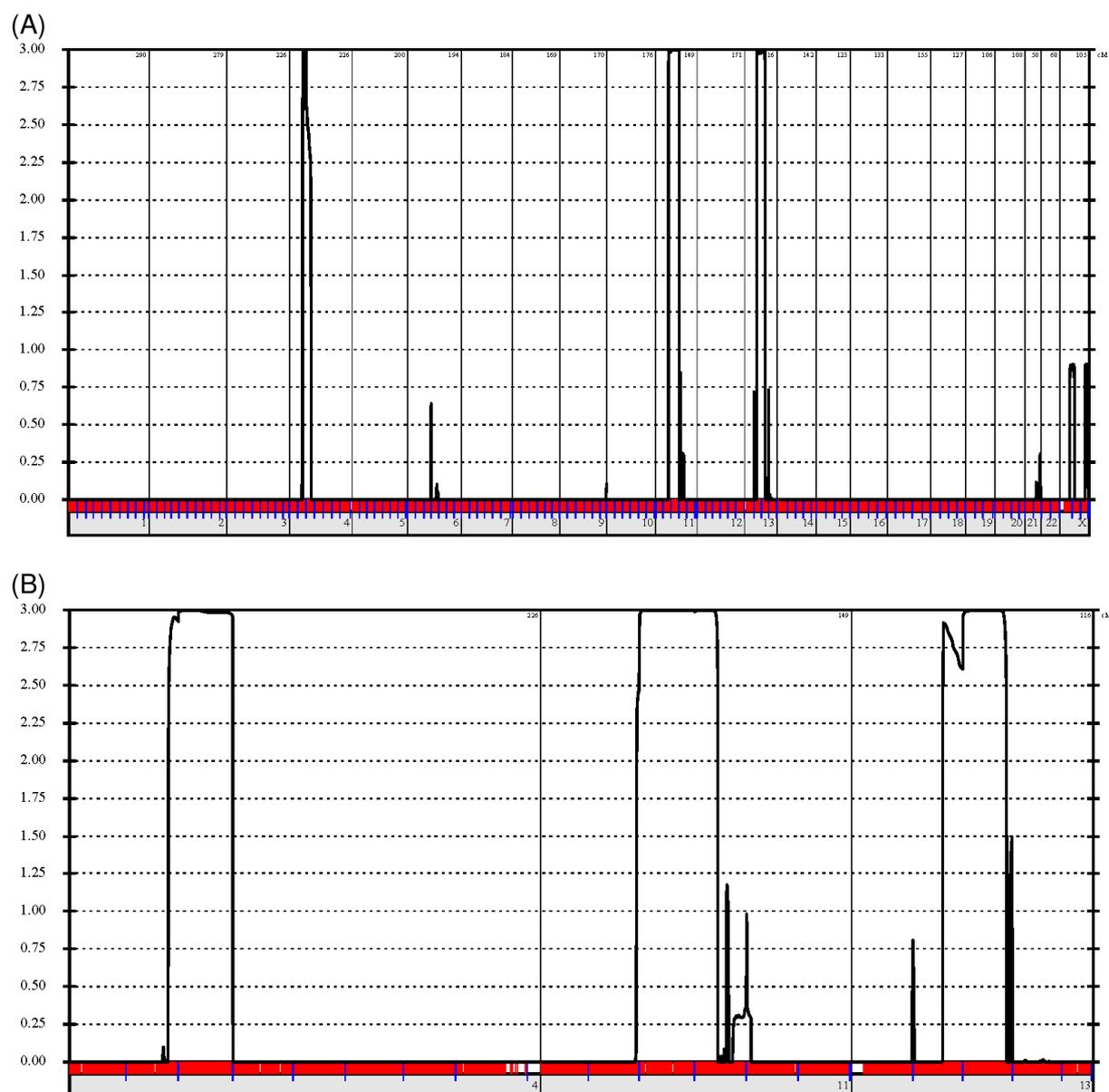
Online Figure I. Parametric LOD scores obtained in family FH4-0102. (A) Multi-point linkage analysis computed in sets of 200 markers with 0.1000 cM spacing, indicating suggestive linkage (LOD 3.0) on chromosome 4p15.1-q13.3, 11p13-q14.1 and 13q14.13-q32.1. (B) Linkage peaks computed in sets of 175 markers with 0.1000 cM spacing corresponded to a 38 Mb region at chromosome 4 (chr. 4: 32,194,465-70,584,037), a 47 Mb region at chromosome 11 (chr. 11: 35,802,039-82,312,013), and a 49 Mb region at chromosome 13 (chr 13: 46,960,154-96,027,407).

Supplementary note

Online Figure II. *STAP1* families

(A)

UPN
Age
TC [pTC]
LDL [pLDL]
HDL [pHDL]
TG [pTG]
Medication
CVD Age onset

I-1
I-2
96
9.41 [99]
7.11 [99]
1.63 [57]
1.49 [69]

CABG 68

II-1
68
8.6 [99]
5.65 [95]
1.6 [50]
3 [99]

II-2
46
8 [99]

II-3

II-4
55
9 [99]
6.65 [99]
1.71 [63]
1.43 [59]

II-5
62
9 [99]

II-6
68
6.99 [83]
4.84 [78]
1.68 [57]
1.04 [30]

II-7

II-8
66
7.47 [99]
5.16 [90]
1.47 [66]
1.86 [76]

II-9
64
5.54 [34]
3.28 [26]
1.54 [47]
1.6 [75]

II-10
62
6.29 [68]
3.66 [41] ([99])
1.45 [38]
2.63 [91]
40mg Lipitor

II-11
60
6.37 [83]
4.48 [81]
.98 [18]
2.03 [87]

II-12
44
8.7 [99]
6.33 [99]
.93 [25]
3.2 [90]

III-1
38
7.37 [99]
4.92 [99]
1.04 [39]
3.13 [91]

III-2
39
7.39 [99]
5.06 [99]
.97 [29]
3.03 [91]

III-3
42
7.47 [99]
4.85 [99]
.97 [30]
3.66 [99]

III-4

III-5
46
6.33 [84]
4.21 [85]
.67 [1]
3.22 [99]

III-6
45
6.73 [92]
4.45 [90]
1.67 [67]
1.36 [67]

III-7
43
6.34 [92]
4.29 [90]
1.58 [65]
1.04 [54]

III-8

III-9
42
5.81 [71]
4.15 [79]
1.12 [51]
1.21 [38]

III-10
41
3.6 [1]
2.02 [7]
1.39 [42]
.42 [1]

III-11
35
5.05 [49]
3.21 [41]
1.31 [80]
1.18 [45]

III-12
35
4.31 [27]
2.59 [30]
1.42 [55]
.66 [23]

IV-1
10
4.53 [80]
2.78 [72]
1.5 [74]
.56 [19]

IV-2
14
4.36 [67]
2.58 [60]
1.44 [55]
.76 [66]

IV-3
12
6.05 [99]
4.17 [99]
1.33 [49]
1.22 [92]

(B)

UPN
Age
TC [pTC]
LDL [pLDL]
HDL [pHDL]
TG [pTG]
Medication
CVD Age onset

I-1
I-2

II-1
79
6.37 [71]
4.13 [65] ([99])
1.27 [27]
2.15 [90]
10mg pravastatin

II-2
63
7.25 [93]
5.34 [95]
.97 [17]
2.09 [88]

II-3
88
5.82 [50]
3.4 [31]
1.18 [21]
2.75 [82]

III-1
53
6.49 [82]
4.44 [83]
1.49 [40]
1.24 [55]

III-2
56
6.49 [86]
4.43 [78]
1.57 [86]
1.09 [34]

III-3
45
6.86 [93]
4.66 [87]
.82 [8]
3.06 [94]

III-4
52
8.3 [99]
5.6 [99]
1.2 [18]
3.2 [99]

III-5

PTCA 43

AP

IV-1
28
6.18 [99]
4.54 [99]
1.08 [41]
1.24 [69]

IV-2
17
6.5 [99]
4.59 [99]
1.33 [51]
1.29 [90]

IV-3
25
5.5 [88]
3.68 [88]
1.55 [66]
.61 [22]

IV-4
28
5.39 [84]
3.49 [84]
1.38 [44]
1.16 [74]

IV-5
26
5.74 [90]
3.63 [78]
1.28 [73]
1.84 [89]

Supplementary note

(C)

UPN
Age
TC [pTC]
LDL [pLDL]
HDL [pHDL]
TG [pTG]
CVD Age onset

I-1   I-2

II-1   II-2

53163   53180
II-3   II-4
85   76
6.29 [82]   5.05 [23]
4.18 [72]   2.84 [11]
1.7 [89]   1.46 [42]
.92 [20]   1.67 [79]
AP

AMI 50

53162   50247   50281   50269
III-1   III-2   III-3   III-4
46   52   53   40
5.83 [71]   7.2 [99]   5.1 [33]   4.66 [26]
4.35 [88]   5.3 [99]   3.56 [54]   3.36 [43]
1.05 [11]   1.4 [83]   1 [8]   .93 [25]
.96 [39]   1.1 [27]   1.2 [52]   .82 [14]
AP   AP 51

50248   34967
IV-1   IV-1
22   17
5.27 [93]   5.65 [99]
4.06 [99]   4.36 [99]
.74 [1]   .86 [1]
1.05 [65]   .95 [75]

(D)

UPN
Age
TC [pTC]
LDL [pLDL]   I-1   I-2
HDL [pHDL]
TG [pTG]

II-1   II-2
45
11 [99]

III-1
23
7.02 [99]
5.16 [99]
1.54 [71]
.72 [30]

Online Figure II. *STAP1* families. Half-blackened symbols indicate affected members, quarter-blackened symbols indicate possibly affected members, and unblackened symbols indicate unaffected members. Age (in years) at lipid measurement, total cholesterol (TC), LDL cholesterol (LDL-c), HDL cholesterol (HDL-c), triglycerides (TG) in mmol/L, history and age onset of coronary artery disease and therapeutic cholesterol-lowering intervention are given. Percentiles matched for age and gender, pTC, pLDL, pHDL, pTG and pLDL corrected for cholesterol-lowering therapy when applicable are given in brackets, a percentile indicated by 99 resembles a percentile above 95th, (a) pedigree FH4-0150 with p.Leu69Ser carriers represented by the dot, (b) pedigree FH4-0247 with p.Ile71Thr carriers represented by the dot, (c) pedigree FH4-0317 with p.Glu97Asp carriers represented by the dot, (d) pedigree FH4-0356 with p.Asp202Asn carriers represented by the dot. The triangles represent the probands. Only branches in which the *STAP1* variant was present are depicted here.
AMI:acute myocardial infarction, AP:angina pectoris, AMI:acute myocardial infarction, CABG:coronary artery bypass graft, PTCA:percutaneous transluminal coronary angioplasty.

Online Figure III. Gene structure of *STAP1.*

PH: Pleckstrin homology domain, SH2:Src homology 2 domain. The positions of tyrosine residues are indicated by arrowheads.

Supplementary note