Use of plasma metabolomics to analyze phenotype-genotype relationships in young hypercholesterolemic females

Zhang, Xiang; Rimbert, Antoine; Balder, Willem; Zwinderman, Aeilko H; Kuivenhoven, Jan Albert; Dallinga-Thie, Geesje M; Groen, Albert K

Published in:
Journal of Lipid Research

DOI:
10.1194/jlr.M088930

Use of Plasma Metabolomics to Analyze Phenotype-Genotype Relationships in Young Hypercholesterolemic Females

Running title: Phenotype-genotype relation in hypercholesterolemic women

Xiang Zhang[1], Antoine Rimbert[2], Willem Balder[2,3], Aeilko Having Zwinderman[4], Jan Albert Kuivenhoven[2], Geesje Margaretha Dallinga-Thie[1,*], Albert Kornelis Groen[1,5,*]

1. Department of Experimental Vascular Medicine, Amsterdam UMC, University of Amsterdam, Meibergdreef 9, Amsterdam, The Netherlands
2. Department of Pediatrics, Section Molecular Genetics, University Medical Center Groningen, University of Groningen, Antonius Deusinglaan 1, Groningen, The Netherlands
3. Department of Cardiology, Jeroen Bosch Hospital, 's-Hertogenbosch, The Netherlands
4. Department of Clinical Epidemiology, Biostatistics, and Bioinformatics, Amsterdam UMC, University of Amsterdam, Meibergdreef 9, Amsterdam, The Netherlands
5. Department of Pediatrics, University Medical Center Groningen, University of Groningen, Antonius Deusinglaan 1, Groningen, The Netherlands

* These authors contributed equally to the study

Abstract

Hypercholesterolemia is characterized by high plasma low density lipoprotein (LDL) cholesterol and often caused by genetic mutations in *LDLR*, *APOB* or *PCSK9*. However, a substantial proportion of hypercholesterolemic subjects do not have any mutations in these canonical genes, leaving the underlying pathobiology to be determined. In this study, we investigated whether combining plasma metabolomics with genetic information increases insight in the biology of hypercholesterolemia. For this proof of concept study, we combined plasma metabolites from 119 hypercholesterolemic females with genetic information on the LDL canonical genes. Using hierarchical clustering we identified four subtypes of hypercholesterolemia, which could be distinguished along two axes represented by triglyceride and large LDL particle concentration. Subjects with mutations in *LDLR* or *APOB* preferentially clustered together suggesting that patients with defects in the LDL receptor pathway show a distinctive metabolomics profile. In conclusion, we show the potential of using metabolomics to segregate hypercholesterolemic subjects in different clusters, which may help in targeting genetic analysis.

Keywords: Hypercholesterolemia; Triglyceride; LDL; Genetics; Metabolomics;

Introduction

Hypercholesterolemia due to a high concentration of plasma low density lipoprotein (LDL) cholesterol has been shown to be a causal factor in accelerating atherosclerosis in a plethora of studies (1, 2). The liver plays a pivotal role in the regulation of cholesterol metabolism. It secretes cholesterol packaged in VLDL (very low density lipoprotein) particles that are subsequently converted into IDL (intermediate density lipoprotein) and LDL particles largely by the action of different lipases in the periphery (3). A key step in the uptake of cholesterol is the internalization of LDL via the LDL receptor (LDLR) (4). Mutations in the LDLR as well as mutations in genes encoding apolipoprotein B (APOB) or proprotein convertase subtilisin/kexin type 9 (PCSK9), are causally related with hypercholesterolemia (5). These genetic mutations, however, do not explain all hypercholesterolemic cases. For instance, in the UK pilot cascade project, 403 of 635 (63.5%) hypercholesterolemic subjects did not have mutations in *LDLR*, *APOB*, or *PCSK9* (6). In a recent large scale study designed to evaluate the prevalence of a familial hypercholesterolemia (FH) mutation among individuals with severe hypercholesterolemia (7), only 24 of 1,386 subjects with LDL cholesterol above 5 mmol/L were identified to have mutations in these three canonical genes. Although the prevalence of genetically defined hypercholesterolemia varies across studies (8), a substantial proportion of hypercholesterolemic subjects do not have mutations in *LDLR*, *APOB* or *PCSK9*. A major reason for this finding will be the presence of disease-causing mutations in other genes involved in cholesterol homeostasis either affecting the LDL receptor pathway or other yet to be defined mechanisms. Interestingly, whole exome sequencing of a cohort with FH subjects without mutations in *LDLR*, *APOB* and *PCSK9* did not identify novel causal mutations (9).

Recently, we analyzed a cohort of 119 young females with plasma LDL cholesterol above the 99th percentile for their age. In 20 hypercholesterolemic females, we identified 12 causal heterozygous mutations in *LDLR* and one causal heterozygous mutation in *APOB* (10). In the 99 remaining females we found 8 subjects carrying a variant in *LDLR* or *APOB* with unknown clinical significance (10). This left us with 91 females that suffered from hypercholesterolemia caused by either a polygenic (11) or epigenetic

(12) mechanism, or presence of pathogenic variant in yet unknown genes. To get further insight in the underlying pathobiology of hypercholesterolemia of unknown origin, we performed plasma metabolite analysis on all the 119 hypercholesterolemic females. We hypothesized that mutations in genes belonging to the same metabolic pathway (e.g. the LDL receptor pathway) should render a similar plasma metabolome. This analysis differentiated four subgroups, which could be distinguished along two axes represented by plasma triglyceride and large LDL particle concentration.

Materials and Methods

Participants

The selection of the participants (N = 119) in this study is described in detail elsewhere (10). In brief, these women were apparently healthy, aged 25 to 40 year and had plasma LDL cholesterol level above 4.7 mmol/l. Exclusion criteria were diagnosis of cardiovascular disease (e.g. myocardial infarction, stroke or coronary surgery), diabetes mellitus, use of lipid-lowering drug, or having aberrant thyroid, liver or kidney function.

Next generation sequencing

With a custom target sequencing array developed based on the SureSelect capture system, we sequenced the coding regions of 11 genes, including *LDLR*, *APOB*, *PCSK9*, *LDLRAP1*, *APOE*, *ABCG5*, *LIPA*, *STAP1*, *MTTP*, *ANGPTL3*, and *SAR1B* to assess a monogenic cause of hypercholesterolemia. If a mutation had minor allele frequency below 0.1% in the Genome of Netherlands (13), it was considered a rare mutation. Mutations that are verified to cause hypercholesterolemia were listed in our previous publication (10).

Detection of copy number variations (CNV) was performed using the CoNVaDING (Copy Number Variation Detection in Next-generation sequencing Gene panels) (14). Detected CNVs were validated using either multiplex ligation-dependent probe amplification, or by long-range PCR or real-time PCR (10).

Genetic risk score calculation

To study a possible polygenic cause of hypercholesterolemia, we calculated the weighted genetic risk score (wGRS). The Global Lipid Genetic Consortium (GLGC) meta-analysis of genome-wide association studies identified 95 loci affecting LDL cholesterol concentration (15). Among these loci, 12 single nucleotide polymorphisms (SNP) had the highest power to discriminate between FH mutation-negative individuals and the general population (11, 16). For each individual, we calculated the wGRS using the

weighted sum of the risk allele (the LDL-C-raising allele) (10). The weights used were the corresponding per-allele effect in plasma LDL cholesterol changes reported by the GLGC (15).

Lifestyle score calculation

To investigate the association between lifestyle and plasma metabolome in hypercholesterolemic females, we used a recently described healthy lifestyle score (17). Points were given for the major lifestyle parameters including smoking status and eating habits. The details were described in our previous publication (10). In short, a maximum of 4 points reflects a very healthy lifestyle. The smaller the score is, the less healthy lifestyle will be. The minimum point is 0.

Metabolite measurements

Fasting plasma samples were routinely collected by Lifelines (www.lifelines.nl) and stored at −80° until analysis on the Nightingale metabolomics platform (Nightingale Health, Finland). This platform includes 225 metabolic features including lipids, lipoproteins, fatty acids, amino acids, and glycolysis precursor molecules listed on https://nightingalehealth.com/biomarkers, using a NMR spectroscopy platform (18, 19).

Statistical analysis

To explore subtypes of hypercholesterolemia, we performed hierarchical clustering based on the plasma metabolomics data. Since the metabolomics data contains measurements of different units, we first scaled the data so that every variable had mean 0 and standard deviation 1. Next, we ran the hierarchical clustering with the function hclust from R. We used Elucidation distance as the dissimilarity measure and complete linkage as the similarity measure between the clusters. The dendrogram was made by using the ggdendro and ggplot2 (20) R package. Finally, we cut the dendrogram into four clusters by using cutree function in R.

To identify the cluster corresponding to hypercholesterolemia due to defects in the LDL receptor pathway, we performed principal component analysis (PCA) on the metabolomics data. Since the data contains measurements of different units, we converted the metabolomics data into ranks, so that every metabolite

had value ranging between 1 and 119. We then calculated the covariance matrix and performed eigenvector decomposition. Entries of every eigenvector are also called loadings. Based on the loadings, we identified metabolites that most correlated to the first and second principal components by calculating the Spearman correlation coefficients.

To evaluate associations between genetic risk/lifestyle scores and metabolite concentrations, we applied a nonparametric method, namely the Kendall's tau correlation test. We reported the Kendall's tau correlation coefficient and p value. A p value below 0.05 is considered significant.

Result

A group of 119 young women with hypercholesterolemia, defined as plasma LDL cholesterol levels above the 99th percentile for their age, was selected from the Lifelines cohort. The baseline characteristics are presented in Table 1. To analyze the underlying pathobiology of the hypercholesterolemic phenotype, plasma metabolomics was performed using the Nightingale platform. Although the absolute values measured in the Nightingale platform are lower than the conventional measured plasma lipids, the two measurements showed similar pattern (Table 1). A summary of all the results of metabolite analysis is presented in Supplemental Table S1. Hierarchical clustering analysis of the metabolomics data set revealed three main clusters and one cluster containing only one sample (Figure 1). The size of the cluster 1, 2, 3, and 4 was 43, 15, 60 and 1, respectively.

To analyze the divergence of the different clusters, we ran principal component analysis. The first and second principal component explained 38% and 21% of the total variance of the metabolic variables across the 119 individuals, respectively (Figure 2). To understand which metabolites correspond to the first and second principal component the most, we calculated the Spearman correlation coefficients between original variables and principal components (Supplemental Table S2). We observed that plasma triglyceride and large LDL particle concentration were the most correlated variables with the PC1 (Spearman correlation coefficient -0.988) and PC2 (Spearman correlation coefficient -0.978), respectively. Therefore, we used these two variables to represent the axes of PC1 and PC2 (Figure 3). Our next question was whether the 4 clusters derived from the hierarchical clustering analysis (Figure 1) were indeed separated by PC1 and PC2. To answer that, we added the hierarchical clustering results to the scatterplot (Figure 3). Inspection reveals that the females in cluster 3 are separated from the other groups by showing a high plasma large LDL particle concentration coupled with relatively low plasma triglyceride, suggesting a defect in hepatic LDL uptake.

Since we sequenced *LDLR*, *APOB* and *PCSK9* in all subjects, we could verify whether indeed the females with known heterozygous mutations in the LDL receptor pathway plot in the region of cluster 3. Indeed,

from 20 subjects with heterozygous mutations in *LDLR* or *APOB* 15 subjects were located in cluster 3 (Figure 4). The other 5 carriers were found in cluster 1 (n = 3) and cluster 2 (n = 2). In addition, we identified 8 women who were heterozygous carrier of a novel variant in *LDLR* or *APOB* from which the pathogenicity has not yet been determined. Five of these 8 subjects were positioned in cluster 3 and three in cluster 1 (Figure 5).

To improve our understanding of the underlying pathobiology of the elevated plasma LDL cholesterol in the remaining 91 women, we calculated the weighted genetic risk score (wGRS) and lifestyle score, and assessed the associations between both scores and plasma concentrations of large LDL particle and triglyceride. As shown in Supplemental Figure S1 and Supplemental Figure S2, no relation could be demonstrated between both scores and plasma large LDL particle concentration (wGRS: Kendall tau correlation coefficient -0.017, p value = 0.80. Lifestyle score: Kendall tau correlation coefficient -0.04, p value = 0.57). Both scores showed moderate association with plasma triglyceride concentration (wGRS: Kendall tau correlation coefficient -0.156, p value = 0.02. Lifestyle score: Kendall tau correlation coefficient -0.198, p value = 0.0099).

Discussion

In the current study, we showed that combining plasma metabolomics data with genetic information can improve our understanding of the origin of severe hypercholesterolemia in young healthy women. These analyses may help the diagnosis and personalized treatment of patients with hypercholesterolemia in which no causal mutations in the canonical LDL genes can be identified.

Metabolic profiling has been used in a large number of cohort studies to assess the value of circulating metabolites in prediction of risk for cardiovascular events (21, 22). More specifically, metabolomics has been used to study associations between circulating metabolites and statin usage (23), CETP inhibition (24) and PCSK9 inhibition (25), generating insight in the broad metabolic effects of these interventions. Nightingale metabolomics data contain not only concentrations in different units, but also other quantities such as ratios, percentages, degrees of saturation and lipoprotein particle size. Therefore, in the current study, we scaled all the metabolic variables to make them have equal importance in the hierarchical clustering.

The hierarchical clustering analysis revealed four clusters in the 119 hypercholesterolemic females with plasma LDL cholesterol above 99th percentile for their age. We hypothesized that mutations in genes belonging to the same metabolic pathway (e.g. the LDL receptor pathway) should render a similar plasma metabolome (one cluster). The principal component analysis revealed that plasma triglyceride and large LDL particle concentrations were the major discriminators for the four clusters. Since cluster 3 is characterized by a high concentration of large LDL particle and relatively low triglyceride in plasma, we hypothesized that this cluster represented the hypercholesterolemia due to defective LDL clearance. Incorporation of the genetic information provided us the verdict, because we expected the 20 subjects carrying a known functional heterozygous mutation in *LDLR* or *APOB* positioning in cluster 3. Indeed, 15 subjects fit this hypothesis and were located in cluster 3.

Then we came up with the question “Can we get insight if a novel variant in *LDLR* or *APOB* is the underlying cause for the severe hypercholesterolemia based on the metabolome profile?”. Indeed, 6 out of

8 carriers of a novel mutation fit in cluster 3, suggesting potential effects of these variants on LDL receptor mediated uptake. This observation suggests that metabolic profiling is useful to delineate the subjects with a pathogenic mutation from those that do not carry any variant in either *LDLR* or *APOB*. However, not all subjects in cluster 3 do carry a variant in *LDLR* or *APOB*. We realize that the pathway of LDL receptor mediated endocytosis and intracellular cholesterol trafficking contains many more genes (26–28) than we have sequenced in our cohort. So expansion of the number of genes on the chip or choosing whole genome sequencing will ultimately improve the information on all genes involved in the LDL receptor pathway and may thus help to identify additional genetic variants underlying the pathobiology in the remaining 40 females in cluster 3. Meanwhile, we cannot exclude other processes underlying the hypercholesterolemia such as epigenetic changes (12), lincRNA (29), microRNA (30) or combinations thereof.

Cluster 4 contained only one subject, and the individual had the highest large LDL particle concentration among the 119 hypercholesterolemic females. Interestingly, we did not identify any mutations in the sequenced genes including *LDLR*, *APOB* and *PCSK9*. This female subject was 28 years with BMI 21.7. Her waist circumference was 69 centimeters. When we compared her plasma metabolomics data to the other 118 hypercholesterolemic females, we identified 77 outlier variables (either below the 1st quantile 1.5×IQR or above the 3rd quantile 1.5×IQR, IQR refers to interquantile range. Supplemental Table S3). We noticed that this female had a high proportion of esterified cholesterol in VLDL and HDL particles compared to the remaining 118 subjects. Interestingly, the CETPtg/apoCI-/- mouse model showed a very similar phenotype (31). Apolipoprotein C1 is an important regulator for CETP activity, which may partly underlie the observed phenotype (32). So far no mutations in *APOC1* have been described.

A recent study (33) showed that hypercholesterolemic subjects without any known genetic defect had lower levels of LDL cholesterol than those with a mutation. Therefore, we hypothesized that the origin of the hypercholesterolemia in cluster 1 may be either polygenic or due to lifestyle factors. Additional analysis of relationships between the wGRS or lifestyle score and triglyceride or large LDL particle

concentration, we observed that only genetic risk scores were negatively associated with triglyceride concentration (Kendall tau correlation coefficient -0.23, p value = 0.04). This observation suggests that this cluster of hypercholesterolemic subjects may be caused by less damaging mutations in genes involved in the LDL receptor pathway. The major observation in the subjects located in cluster 2 is that they had elevated plasma triglyceride. The genetic array used in the current study does not contain the genes involved in triglyceride metabolism. Our data suggest that generation of a triglyceride specific gene array may generate interesting results in the subjects in this cluster.

In summary, this study shows that bioinformatic analysis of metabolomics data derived from hypercholesterolemic subjects generates interesting clusters of patients that may help guiding targeted genomics approaches hypercholesterolemia.

Acknowledgments

We would like to thank all participants of the Lifelines study. This study makes use of data generated by the Genome of the Netherlands Project. A full list of the investigators is available from www.nlgenome.nl. This work was supported by the Netherlands CardioVascular Research Initiative: "the Dutch Heart Foundation, Dutch Federation of University Medical Centers, the Netherlands Organization for Health Research and Development and the Royal Netherlands Academy of Sciences" (CVON2011-2016; Acronym Genius1 and CVON2017-2020; Acronym Genius2 to Dr Kuivenhoven), the European Union (FP7-603091; Acronym TransCard to Dr Kuivenhoven). Dr Kuivenhoven is Established Investigator of the Netherlands Heart Foundation (2015T068).

Reference

1. Ference, B. A., H. N. Ginsberg, I. Graham, K. K. Ray, C. J. Packard, E. Bruckert, R. A. Hegele, R. M. Krauss, F. J. Raal, H. Schunkert, G. F. Watts, J. Borén, S. Fazio, J. D. Horton, L. Masana, S. J. Nicholls, B. G. Nordestgaard, B. van de Sluis, M.-R. Taskinen, L. Tokgözoglu, U. Landmesser, U. Laufs, O. Wiklund, J. K. Stock, M. J. Chapman, and A. L. Catapano. 2017. Low-density lipoproteins cause atherosclerotic cardiovascular disease. 1. evidence from genetic, epidemiologic, and clinical studies. a consensus statement from the european atherosclerosis society consensus panel. Eur Heart J. 38: 2459–2472.

2. Goldstein, J. L., and M. S. Brown. 2015. A century of cholesterol and coronaries: From plaques to genes to statins. Cell. 161: 161–172.

3. Packard, C. J., and J. Shepherd. 1997. Lipoprotein heterogeneity and apolipoprotein b metabolism. Arterioscler Thromb Vasc Biol. 17: 3542–56.

4. Brown, M. S., and J. L. Goldstein. 1986. A receptor-mediated pathway for cholesterol homeostasis. Science (New York, N.Y.). 232: 34–47.

5. Soutar, A. K., and R. P. Naoumova. 2007. Mechanisms of disease: Genetic causes of familial hypercholesterolemia. Nature clinical practice. Cardiovascular medicine. 4: 214–225.

6. Taylor, A., D. Wang, K. Patel, R. Whittall, G. Wood, M. Farrer, R. D. G. Neely, S. Fairgrieve, D. Nair, M. Barbir, J. L. Jones, S. Egan, R. Everdale, Y. Lolin, E. Hughes, J. A. Cooper, S. G. Hadfield, G. Norbury, and S. E. Humphries. 2010. Mutation detection rate and spectrum in familial hypercholesterolaemia patients in the uk pilot cascade project. Clinical genetics. 77: 572–580.

7. Khera, A. V., H.-H. Won, G. M. Peloso, K. S. Lawson, T. M. Bartz, X. Deng, E. M. van Leeuwen, P. Natarajan, C. A. Emdin, A. G. Bick, A. C. Morrison, J. A. Brody, N. Gupta, A. Nomura, T. Kessler, S. Duga, J. C. Bis, C. M. van Duijn, L. A. Cupples, B. Psaty, D. J. Rader, J. Danesh, H. Schunkert, R. McPherson, M. Farrall, H. Watkins, E. Lander, J. G. Wilson, A. Correa, E. Boerwinkle, P. A. Merlini, D.

Ardissino, D. Saleheen, S. Gabriel, and S. Kathiresan. 2016. Diagnostic yield and clinical utility of sequencing familial hypercholesterolemia genes in patients with severe hypercholesterolemia. J Am Coll Cardiol. 67: 2578–89.

8. Wang, J., J. S. Dron, M. R. Ban, J. F. Robinson, A. D. McIntyre, M. Alazzam, P. J. Zhao, A. A. Dilliott, H. Cao, M. W. Huff, D. Rhainds, C. Low-Kam, M.-P. Dubé, G. Lettre, J.-C. Tardif, and R. A. Hegele. 2016. Polygenic versus monogenic causes of hypercholesterolemia ascertained clinically. Arterioscler Thromb Vasc Biol. 36: 2439–2445.

9. Futema, M., V. Plagnol, K. Li, R. A. Whittall, H. A. W. Neil, M. Seed, S. B. Consortium, S. Bertolini, S. Calandra, O. S. Descamps, C. A. Graham, R. A. Hegele, F. Karpe, R. Durst, E. Leitersdorf, N. Lench, D. R. Nair, H. Soran, F. M. Van Bockxmeer, U. Consortium, and S. E. Humphries. 2014. Whole exome sequencing of familial hypercholesterolaemia patients negative for ldlr/apob/pcsk9 mutations. Journal of medical genetics. 51: 537–544.

10. Balder, J.-W., A. Rimbert, X. Zhang, M. Viel, R. Kanninga, F. van Dijk, P. Lansberg, R. Sinke, and J. A. Kuivenhoven. 2018. Genetics, lifestyle, and low-density lipoprotein cholesterol in young and apparently healthy women. Circulation. 137: 820–831.

11. Talmud, P. J., S. Shah, R. Whittall, M. Futema, P. Howard, J. A. Cooper, S. C. Harrison, K. Li, F. Drenos, F. Karpe, H. A. W. Neil, O. S. Descamps, C. Langenberg, N. Lench, M. Kivimaki, J. Whittaker, A. D. Hingorani, M. Kumari, and S. E. Humphries. 2013. Use of low-density lipoprotein cholesterol gene score to distinguish patients with polygenic and monogenic familial hypercholesterolaemia: A case-control study. Lancet (London, England). 381: 1293–1301.

12. Dekkers, K. F., M. van Iterson, R. C. Slieker, M. H. Moed, M. J. Bonder, M. van Galen, H. Mei, D. V. Zhernakova, L. H. van den Berg, J. Deelen, J. van Dongen, D. van Heemst, A. Hofman, J. J. Hottenga, C. J. H. van der Kallen, C. G. Schalkwijk, C. D. A. Stehouwer, E. F. Tigchelaar, A. G. Uitterlinden, G. Willemsen, A. Zhernakova, L. Franke, P. A. C. 't Hoen, R. Jansen, J. van Meurs, D. I. Boomsma, C. M.

van Duijn, M. M. J. van Greevenbroek, J. H. Veldink, C. Wijmenga, BIOS Consortium, E. W. van Zwet, P. E. Slagboom, J. W. Jukema, and B. T. Heijmans. 2016. Blood lipids influence dna methylation in circulating cells. Genome Biol. 17: 138.

13. Netherlands Consortium, G. of the. 2014. Whole-genome sequence variation, population structure and demographic history of the dutch population. Nature genetics. 46: 818–825.

14. Johansson, L. F., F. van Dijk, E. N. de Boer, K. K. van Dijk-Bos, J. D. H. Jongbloed, A. H. van der Hout, H. Westers, R. J. Sinke, M. A. Swertz, R. H. Sijmons, and B. Sikkema-Raddatz. 2016. CoNVaDING: Single exon variation detection in targeted ngs data. Hum Mutat. 37: 457–64.

15. Teslovich, T. M., K. Musunuru, A. V. Smith, A. C. Edmondson, I. M. Stylianou, M. Koseki, J. P. Pirruccello, S. Ripatti, D. I. Chasman, C. J. Willer, C. T. Johansen, S. W. Fouchier, A. Isaacs, G. M. Peloso, M. Barbalic, S. L. Ricketts, J. C. Bis, Y. S. Aulchenko, G. Thorleifsson, M. F. Feitosa, J. Chambers, M. Orho-Melander, O. Melander, T. Johnson, X. Li, X. Guo, M. Li, Y. Shin Cho, M. Jin Go, Y. Jin Kim, J.-Y. Lee, T. Park, K. Kim, X. Sim, R. Twee-Hee Ong, D. C. Croteau-Chonka, L. A. Lange, J. D. Smith, K. Song, J. Hua Zhao, X. Yuan, J. Luan, C. Lamina, A. Ziegler, W. Zhang, R. Y. L. Zee, A. F. Wright, J. C. M. Witteman, J. F. Wilson, G. Willemsen, et al. 2010. Biological, clinical and population relevance of 95 loci for blood lipids. Nature. 466: 707–13.

16. Talmud, P. J., F. Drenos, S. Shah, T. Shah, J. Palmen, C. Verzilli, T. R. Gaunt, J. Pallas, R. Lovering, K. Li, J. P. Casas, R. Sofat, M. Kumari, S. Rodriguez, T. Johnson, S. J. Newhouse, A. Dominiczak, N. J. Samani, M. Caulfield, P. Sever, A. Stanton, D. C. Shields, S. Padmanabhan, O. Melander, C. Hastie, C. Delles, S. Ebrahim, M. G. Marmot, G. D. Smith, D. A. Lawlor, P. B. Munroe, I. N. Day, M. Kivimaki, J. Whittaker, S. E. Humphries, A. D. Hingorani, ASCOT investigators, NORDIL investigators, and BRIGHT Consortium. 2009. Gene-centric association signals for lipids and apolipoproteins identified via the humancvd beadchip. Am J Hum Genet. 85: 628–42.

17. Khera, A. V., C. A. Emdin, I. Drake, P. Natarajan, A. G. Bick, N. R. Cook, D. I. Chasman, U. Baber, R. Mehran, D. J. Rader, V. Fuster, E. Boerwinkle, O. Melander, M. Orho-Melander, P. M. Ridker, and S. Kathiresan. 2016. Genetic risk, adherence to a healthy lifestyle, and coronary disease. N Engl J Med. 375: 2349–2358.

18. Fischer, K., J. Kettunen, P. Würtz, T. Haller, A. S. Havulinna, A. J. Kangas, P. Soininen, T. Esko, M.-L. Tammesoo, R. Mägi, S. Smit, A. Palotie, S. Ripatti, V. Salomaa, M. Ala-Korpela, M. Perola, and A. Metspalu. 2014. Biomarker profiling by nuclear magnetic resonance spectroscopy for the prediction of all-cause mortality: An observational study of 17,345 persons. PLoS Med. 11: e1001606.

19. Soininen, P., A. J. Kangas, P. Würtz, T. Suna, and M. Ala-Korpela. 2015. Quantitative serum nuclear magnetic resonance metabolomics in cardiovascular epidemiology and genetics. Circ Cardiovasc Genet. 8: 192–206.

20. Wickham, H. 2016. Ggplot2: Elegant graphics for data analysis. Springer-Verlag New York.

21. Würtz, P., A. S. Havulinna, P. Soininen, T. Tynkkynen, D. Prieto-Merino, T. Tillin, A. Ghorbani, A. Artati, Q. Wang, M. Tiainen, A. J. Kangas, J. Kettunen, J. Kaikkonen, V. Mikkilä, A. Jula, M. Kähönen, T. Lehtimäki, D. A. Lawlor, T. R. Gaunt, A. D. Hughes, N. Sattar, T. Illig, J. Adamski, T. J. Wang, M. Perola, S. Ripatti, R. S. Vasan, O. T. Raitakari, R. E. Gerszten, J.-P. Casas, N. Chaturvedi, M. Ala-Korpela, and V. Salomaa. 2015. Metabolite profiling and cardiovascular event risk: A prospective study of 3 population-based cohorts. Circulation. 131: 774–85.

22. Holmes, M. V., I. Y. Millwood, C. Kartsonaki, M. R. Hill, D. A. Bennett, R. Boxall, Y. Guo, X. Xu, Z. Bian, R. Hu, R. G. Walters, J. Chen, M. Ala-Korpela, S. Parish, R. J. Clarke, R. Peto, R. Collins, L. Li, Z. Chen, and China Kadoorie Biobank Collaborative Group. 2018. Lipids, lipoproteins, and metabolites and risk of myocardial infarction and stroke. J Am Coll Cardiol. 71: 620–632.

23. Würtz, P., Q. Wang, P. Soininen, A. J. Kangas, G. Fatemifar, T. Tynkkynen, M. Tiainen, M. Perola, T. Tillin, A. D. Hughes, P. Mäntyselkä, M. Kähönen, T. Lehtimäki, N. Sattar, A. D. Hingorani, J.-P. Casas,

V. Salomaa, M. Kivimäki, M.-R. Järvelin, G. Davey Smith, M. Vanhala, D. A. Lawlor, O. T. Raitakari, N. Chaturvedi, J. Kettunen, and M. Ala-Korpela. 2016. Metabolomic profiling of statin use and genetic inhibition of hmg-coa reductase. J Am Coll Cardiol. 67: 1200–1210.

24. Kettunen, J., M. V. Holmes, E. Allara, O. Anufrieva, P. Ohukainen, C. Oliver-Williams, T. Tillin, A. Hughes, M. Kahonen, T. Lehtimaki, J. Viikari, O. Raitakari, V. Salomaa, M.-R. Jarvinen, M. Perola, G. Davey Smith, N. Chaturvedi, J. Danesh, E. Di Angelantonio, A. Butterworth, and M. Ala-Korpela. 2018. Lipoprotein signatures of cholesteryl ester transfer protein and hmg-coa reductase inhibition. bioRxiv. [online] https://www.biorxiv.org/content/early/2018/04/05/295394.

25. Sliz, E., J. Kettunen, M. V. Holmes, C. Oliver-Williams, C. Boachie, Q. Wang, M. Mannikko, S. Sebert, R. Walters, K. Lin, I. Y. Millwood, R. Clarke, L. Lee, N. Rankin, P. Welsh, C. Delles, I. Ford, J. W. Jukema, S. Trompet, M. Perola, V. Salomaa, M.-R. Jarvelin, Z. Chen, D. A. Lawlor, M. Ala-Korpela, J. Danesh, G. Davey Smith, N. Sattar, A. Butterworth, and P. Wurtz. 2018. Metabolomic consequences of genetic inhibition of pcsk9 compared with statin treatment. bioRxiv. [online] https://www.biorxiv.org/content/early/2018/03/14/278861.

26. Marques-Pinheiro, A., M. Marduel, J.-P. Rabès, M. Devillers, L. Villéger, D. Allard, J. Weissenbach, M. Guerin, Y. Zair, D. Erlich, C. Junien, A. Munnich, M. Krempf, M. Abifadel, J.-P. Jaïs, French Research Network on ADH, C. Boileau, and M. Varret. 2010. A fourth locus for autosomal dominant hypercholesterolemia maps at 16q22.1. Eur J Hum Genet. 18: 1236–42.

27. Bartuzi, P., D. D. Billadeau, R. Favier, S. Rong, D. Dekker, A. Fedoseienko, H. Fieten, M. Wijers, J. H. Levels, N. Huijkman, N. Kloosterhuis, H. van der Molen, G. Brufau, A. K. Groen, A. M. Elliott, J. A. Kuivenhoven, B. Plecko, G. Grangl, J. McGaughran, J. D. Horton, E. Burstein, M. H. Hofker, and B. van de Sluis. 2016. CCC- and wash-mediated endosomal sorting of ldlr is required for normal clearance of circulating ldl. Nat Commun. 7: 10961.

28. Paththinige, C. S., N. D. Sirisena, and V. Dissanayake. 2017. Genetic determinants of inherited susceptibility to hypercholesterolemia - a comprehensive literature review. Lipids Health Dis. 16: 103.

29. Hu, Y.-W., J.-Y. Yang, X. Ma, Z.-P. Chen, Y.-R. Hu, J.-Y. Zhao, S.-F. Li, Y.-R. Qiu, J.-B. Lu, Y.-C. Wang, J.-J. Gao, Y.-H. Sha, L. Zheng, and Q. Wang. 2014. A lincRNA-dynlrb2-2/gpr119/glp-1R/abca1-dependent signal transduction pathway is essential for the regulation of cholesterol homeostasis. J Lipid Res. 55: 681–97.

30. Irani, S., J. Iqbal, W. J. Antoni, L. Ijaz, and M. M. Hussain. 2018. MicroRNA-30c reduces plasma cholesterol in homozygous familial hypercholesterolemic and type 2 diabetic mouse models. J Lipid Res. 59: 144–154.

31. Gautier, T., D. Masson, M. C. Jong, L. Duverneuil, N. Le Guern, V. Deckert, J.-P. Pais de Barros, L. Dumont, A. Bataille, Z. Zak, X.-C. Jiang, A. R. Tall, L. M. Havekes, and L. Lagrost. 2002. Apolipoprotein ci deficiency markedly augments plasma lipoprotein changes mediated by human cholesteryl ester transfer protein (cetp) in cetp transgenic/apoci-knocked out mice. J Biol Chem. 277: 31354–63.

32. Pillois, X., T. Gautier, B. Bouillet, J.-P. Pais de Barros, A. Jeannin, B. Vergès, J. Bonnet, and L. Lagrost. 2012. Constitutive inhibition of plasma cetp by apolipoprotein c1 is blunted in dyslipidemic patients with coronary artery disease. J Lipid Res. 53: 1200–9.

33. Lorenzo, A. D., J. D. L. da Silva, C. E. James, A. C. Pereira, and A. S. B. Moreira. 2018. Clinical, anthropometric and biochemical characteristics of patients with or without genetically confirmed familial hypercholesterolemia. Arq Bras Cardiol. 110: 119–123.

Tables

Table 1 Characteristics of 119 hypercholesterolemic females

	N	Mean Standard deviation		
Age (year)	119	32.90 ± 4.37		
BMI	119	27.9 ± 5.10		
		Clinical chemistry	Nightingale metabolomics	Spearman correlation coefficients
LDL cholesterol (mmol/l)	119	5.25 ± 0.50	2.27 ± 0.26	0.66
Total cholesterol (mmol/l)	119	7.17 ± 0.64	5.57 ± 0.43	0.68
Triglyceride (mmol/l)	119	1.50 ± 0.68	1.45 ± 0.47	0.96
HDL cholesterol (mmol/l)	119	1.39 ± 0.28	1.47 ± 0.22	0.84
ApoB (g/l)	119	1.25 ± 0.14	1.10 ± 0.11	0.78

Figures and figure legends

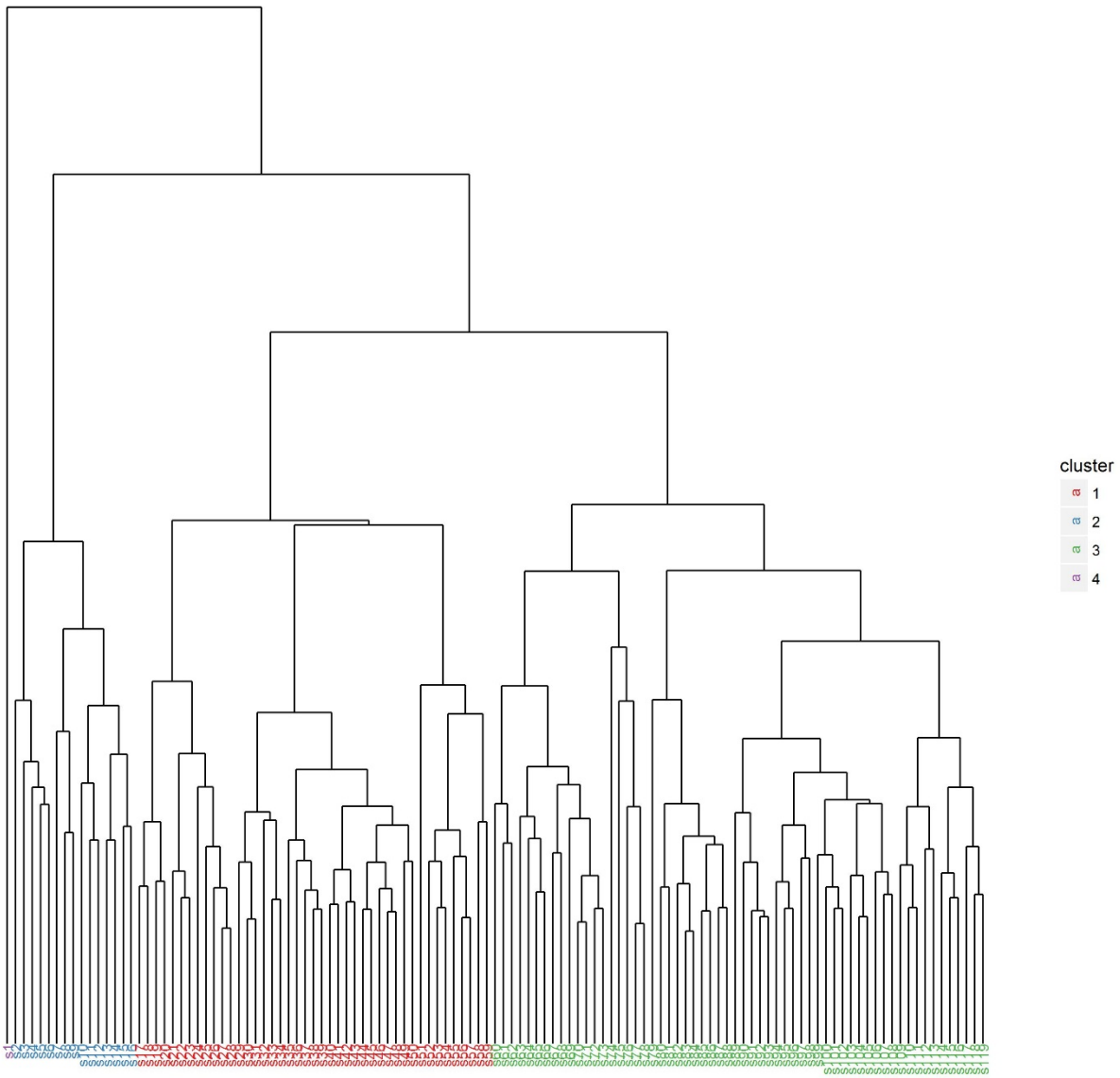

Figure 1 Hierarchical clustering of plasma metabolomics data derived from 119 hypercholesterolemic females. Elucidation distance was used as the dissimilarity measure and complete linkage was used as the dissimilarity measure between the clusters.

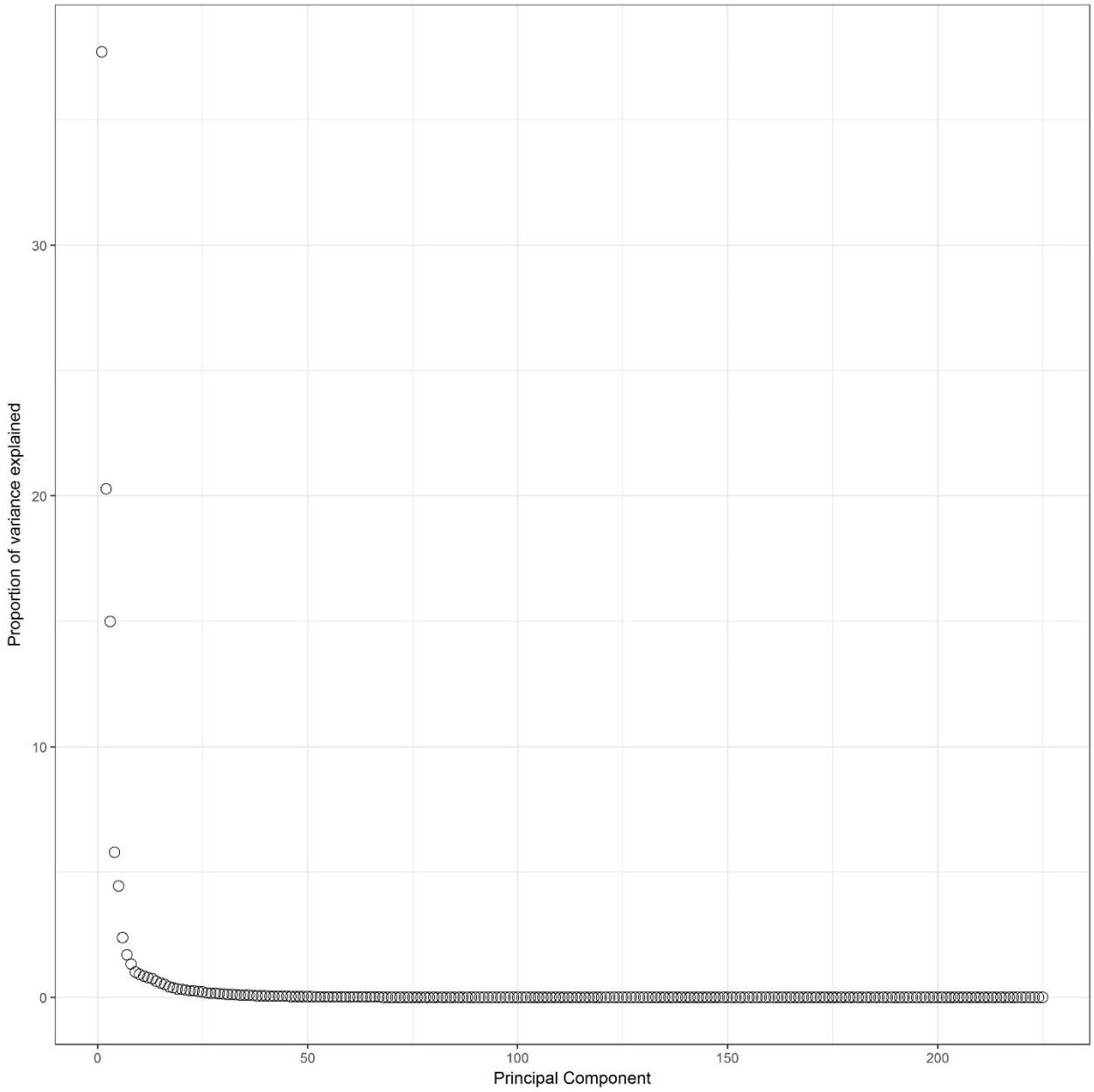

Figure 2 Proportion of variance explained by principal components derived from plasma metabolomics data of 119 hypercholesterolemic females.

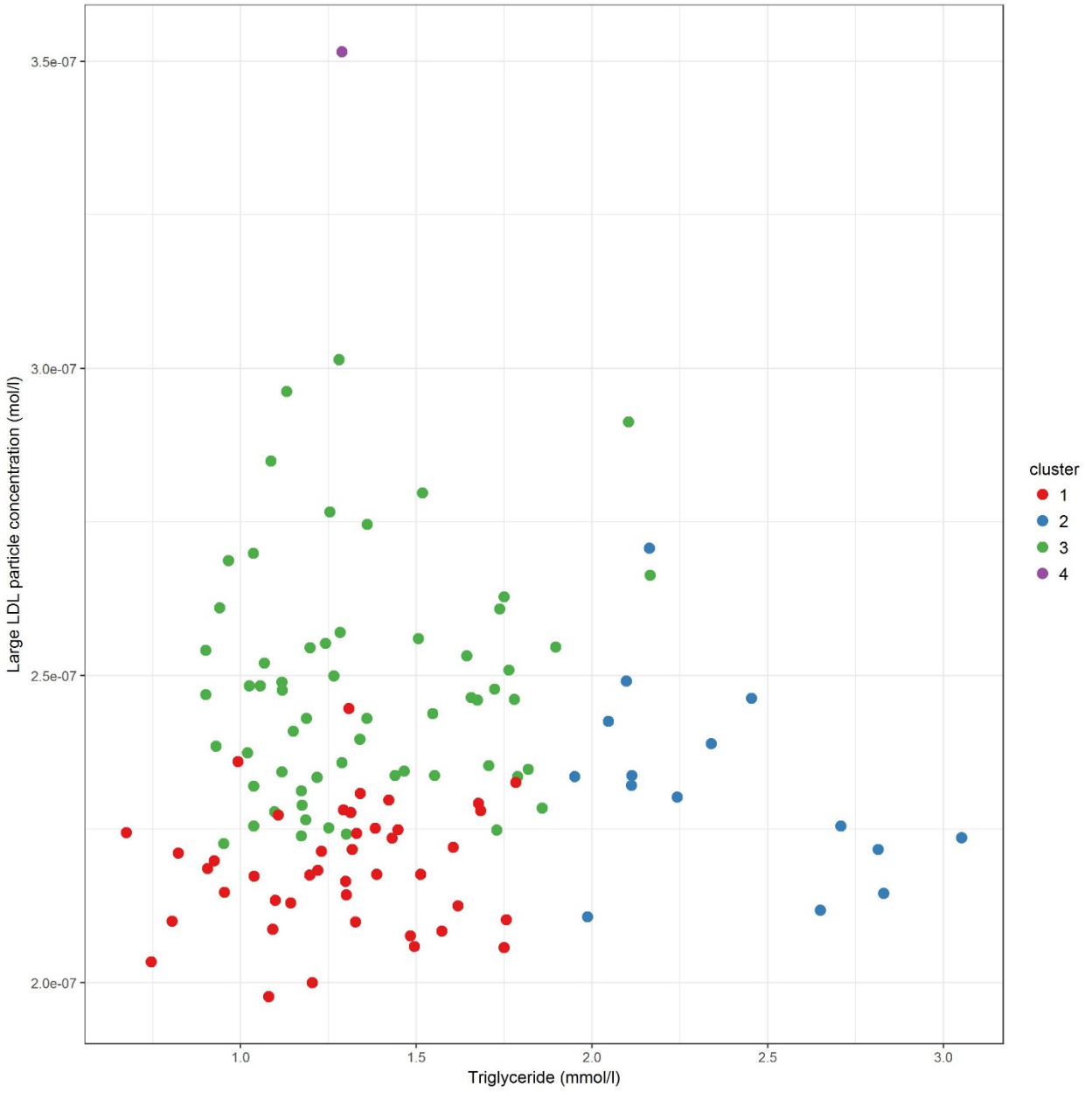

Figure 3 Plasma triglyceride against large LDL particle concentration in 119 hypercholesterolemic females. Different colors refer to the hierarchical clustering outcomes (red: cluster 1, blue: cluster 2, green: cluster 3, purple: cluster 4).

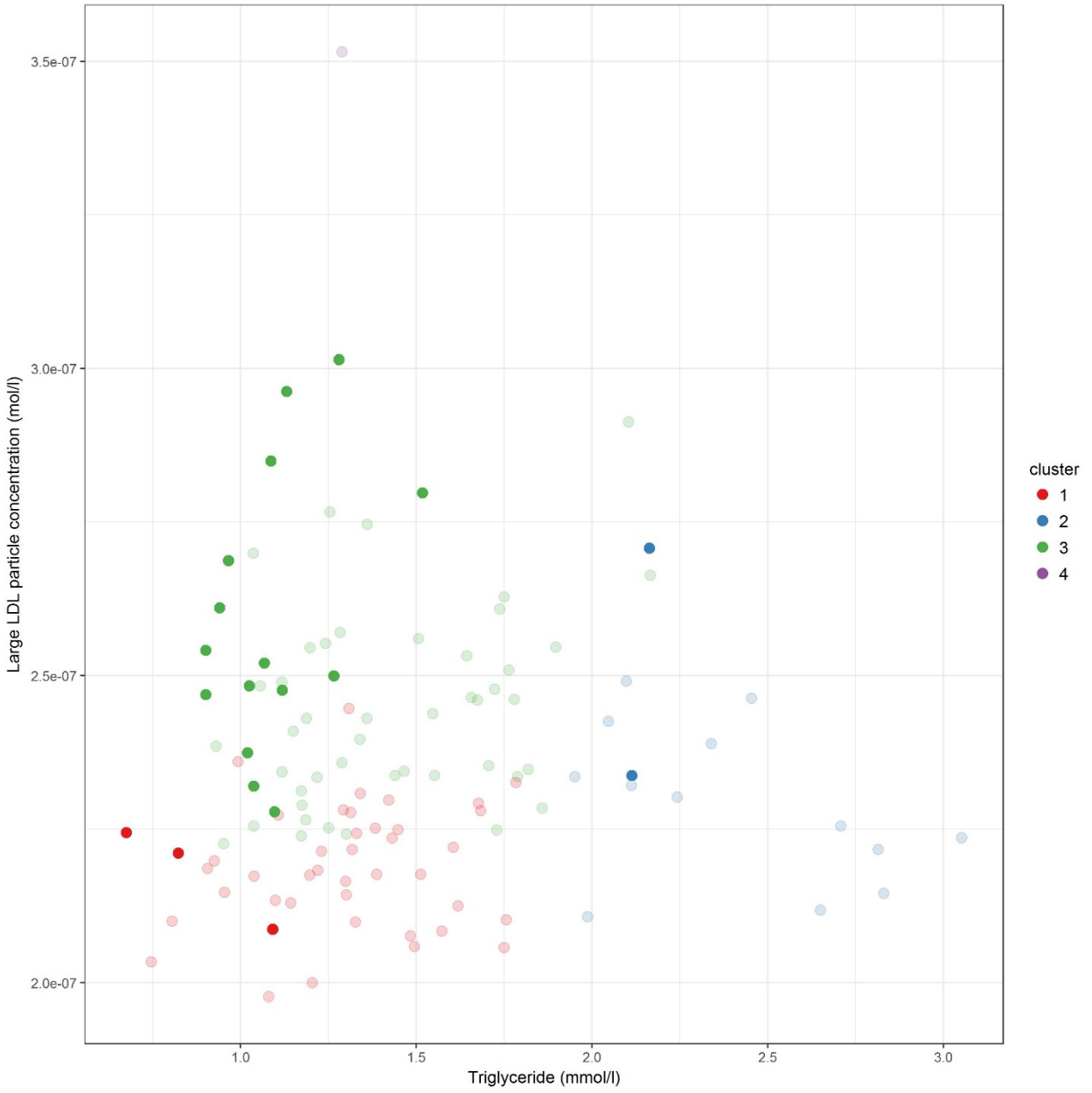

Figure 4 Plasma triglyceride against large LDL particle concentration in 119 hypercholesterolemic females. Different colors refer to the hierarchical clustering outcomes (red: cluster 1, blue: cluster 2, green: cluster 3, purple: cluster 4). The hypercholesterolemic females with mutations that were known to affect the LDL receptor pathway were highlighted.

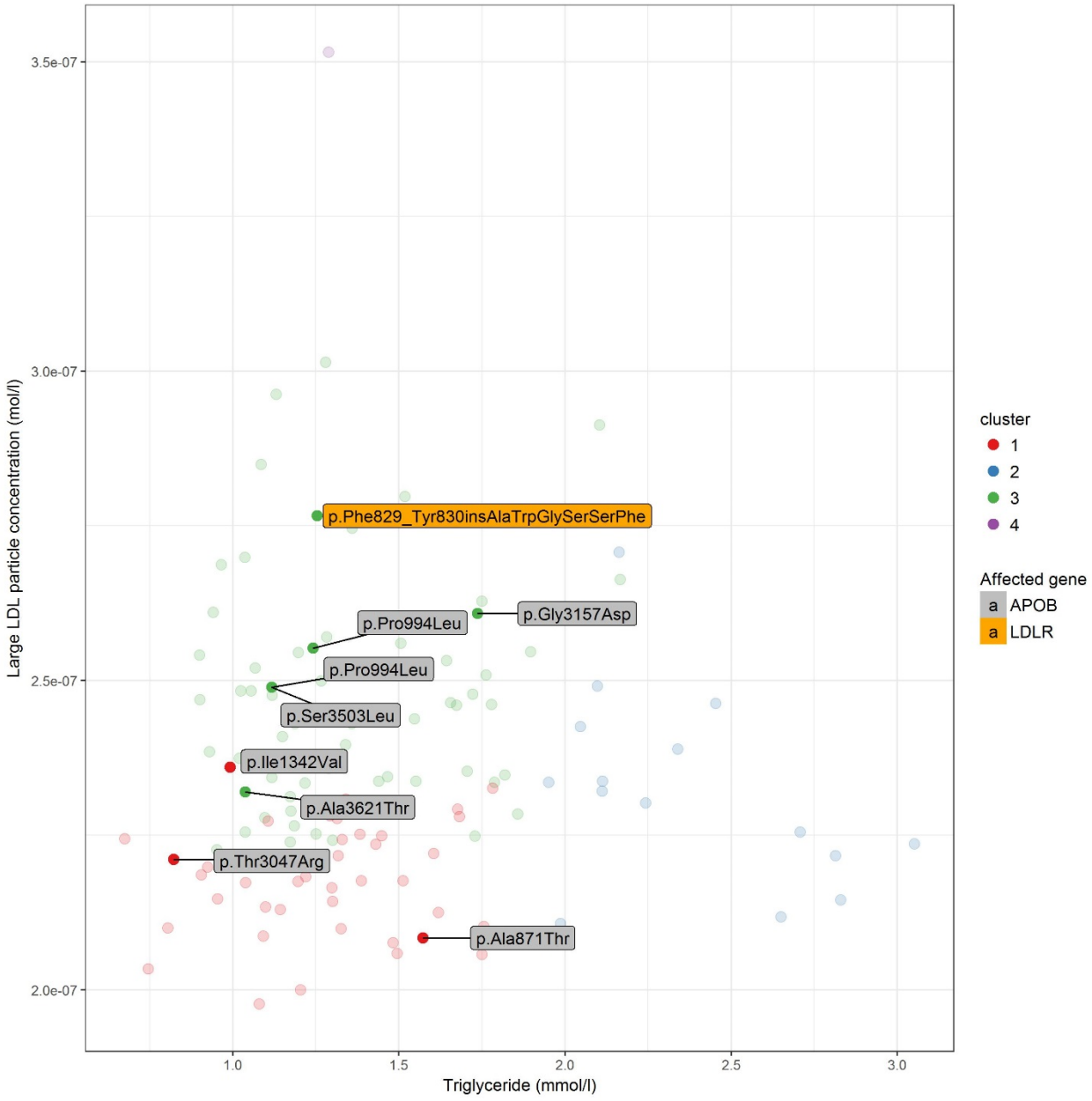

Figure 5 Plasma triglyceride against large LDL particle concentration in 119 hypercholesterolemic females. Different colors refer to the hierarchical clustering outcomes (red: cluster 1, blue: cluster 2, green: cluster 3, purple: cluster 4). The highlighted dots represent 8 individuals who carry a heterozygous variant in *LDLR* or *APOB* of unknown clinical significance. The specific variant in *LDLR* or *APOB* is shown.