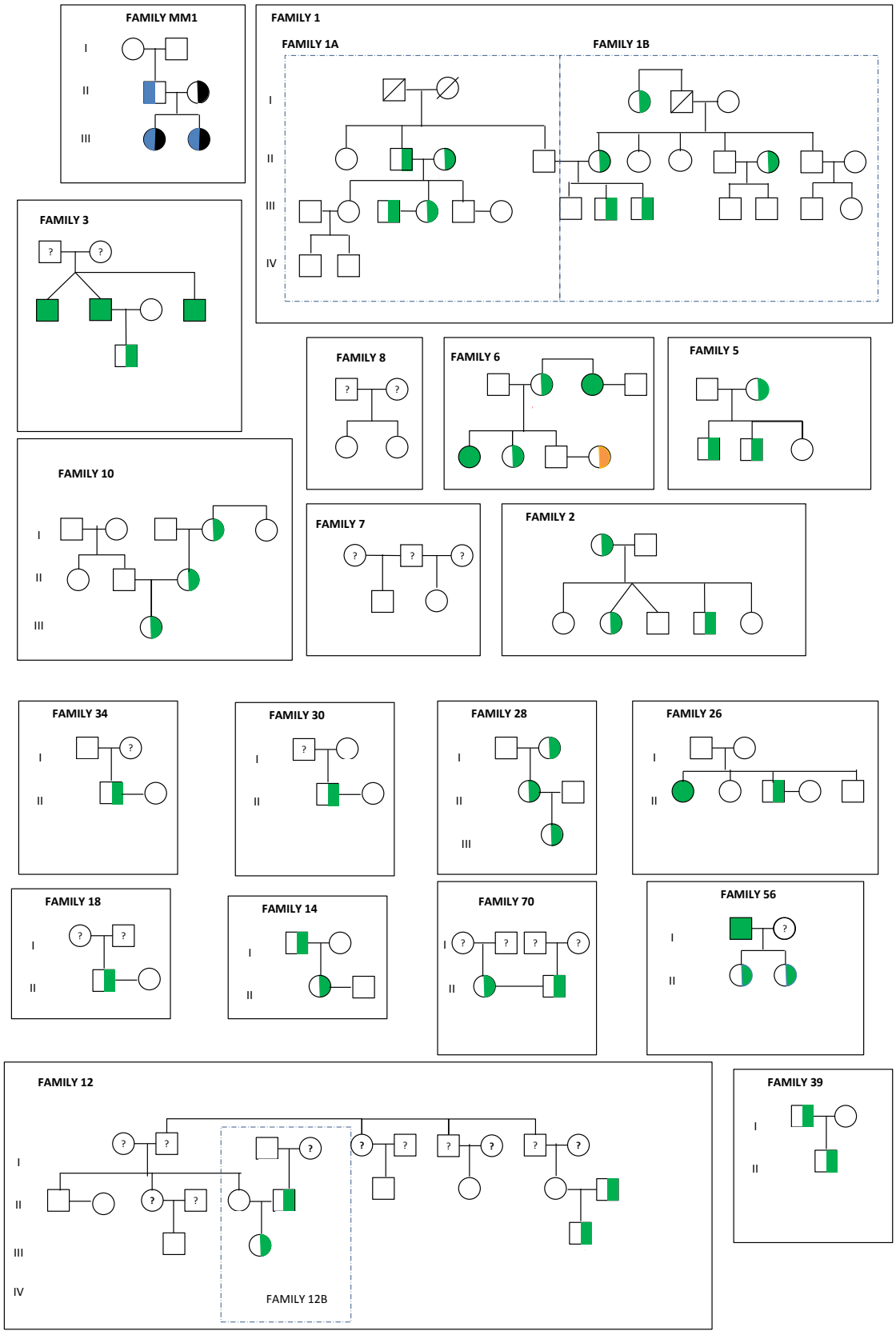
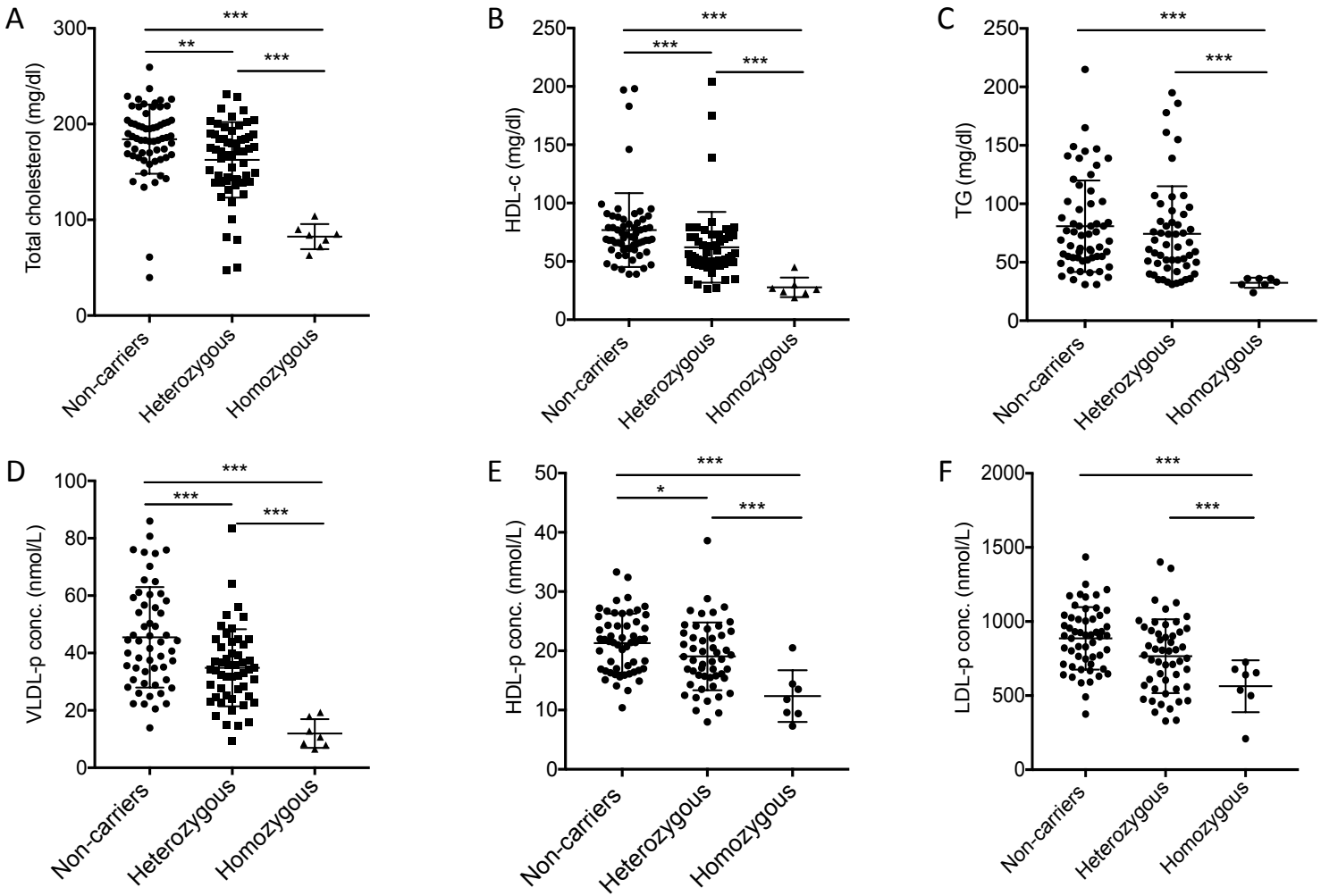
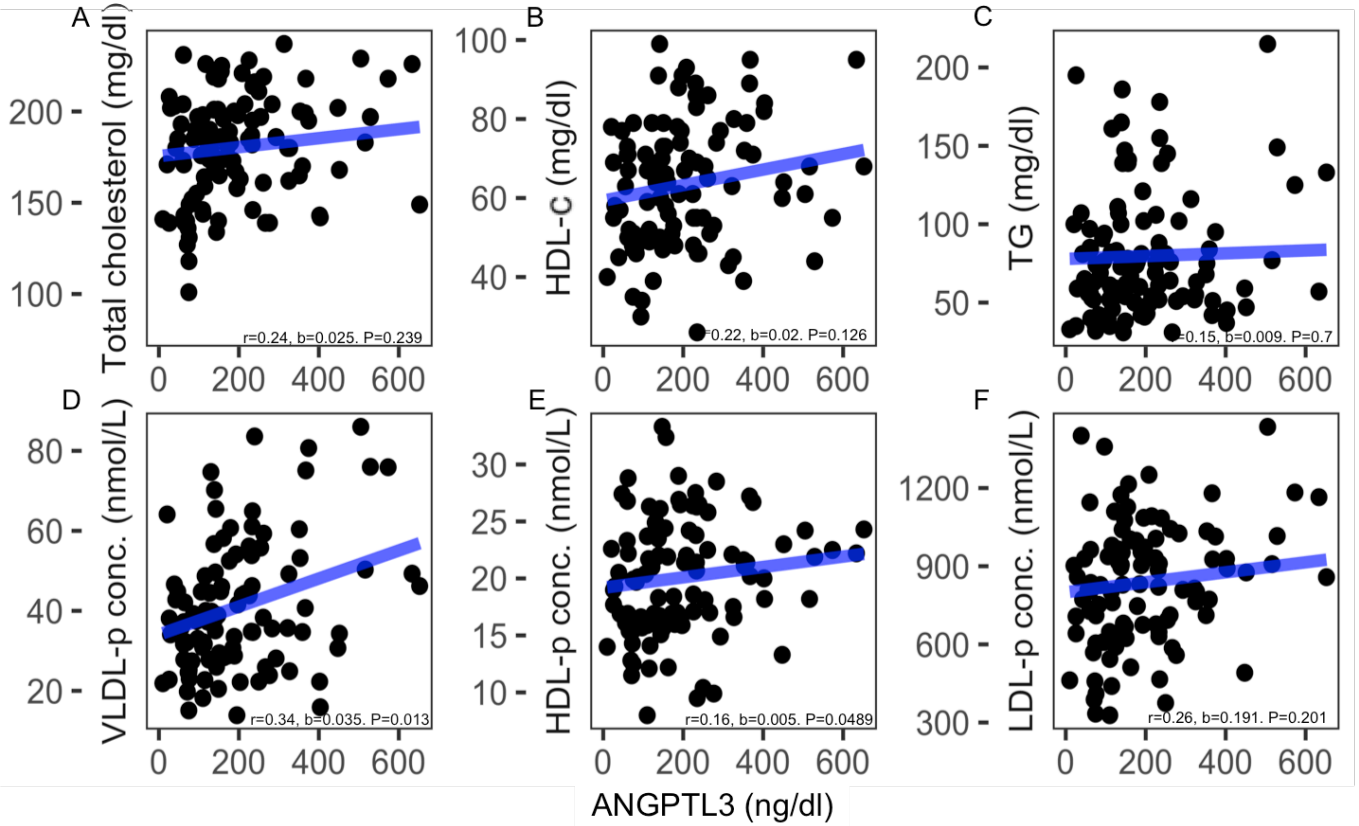
Supplementary data



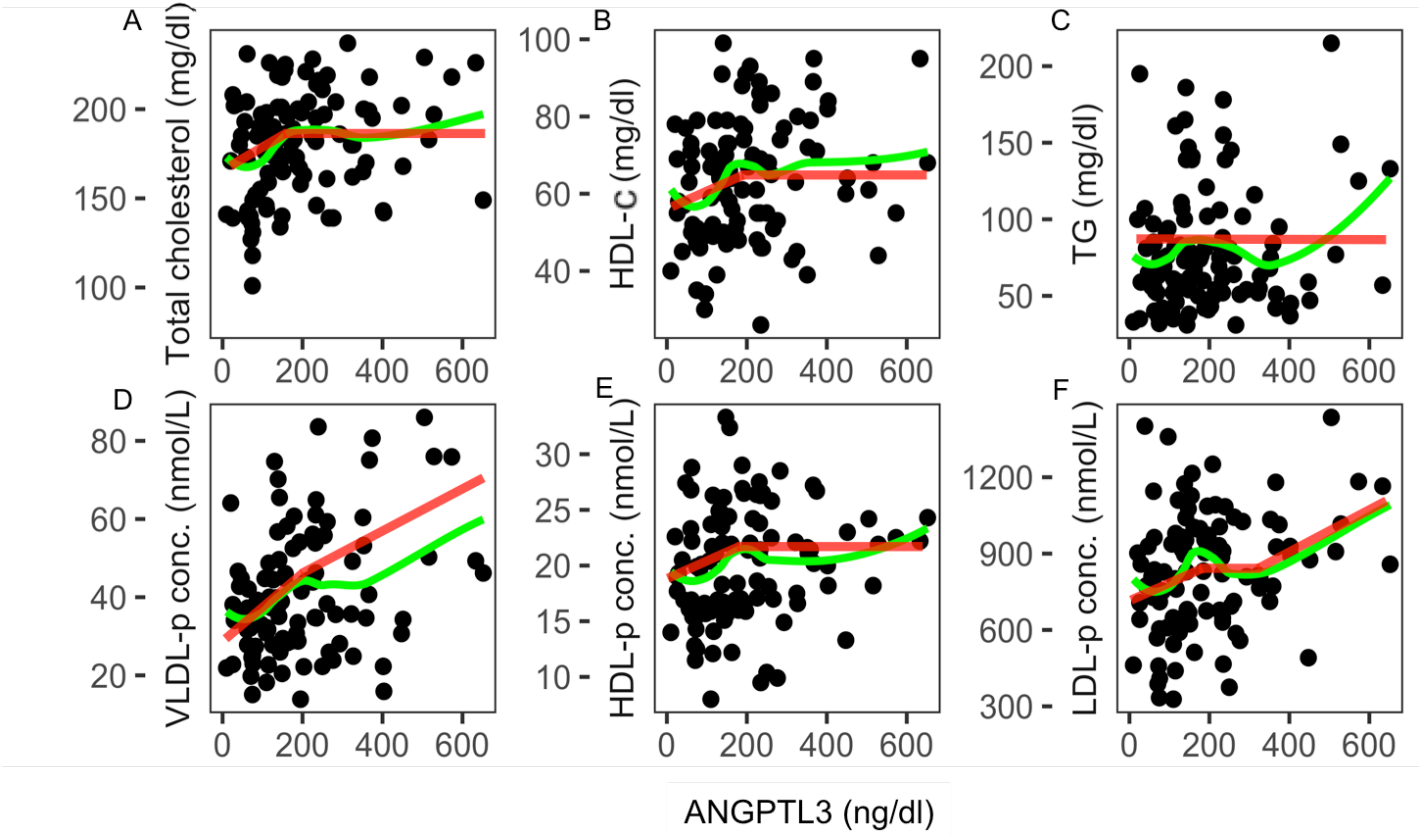
Supplementary Figure 1: Family trees of the 19 families in the study cohort. Colored = carrier of ANGPTL3 loss-of-function mutant. ?=unknown mutation status (not included in the study cohort). Samples were collected from all 127 subjects with known ANGPTL3 mutation status.



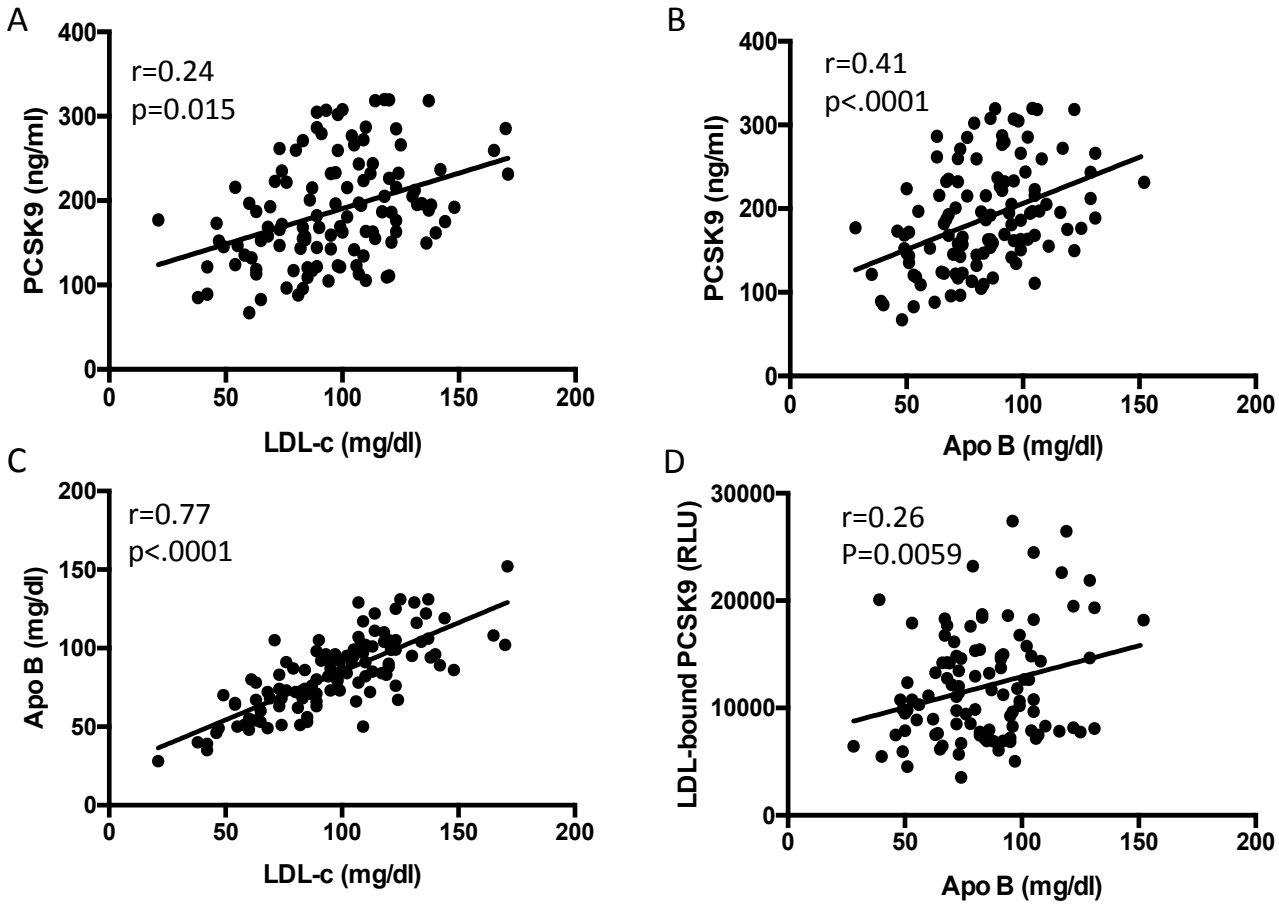
Supplementary Figure 2: Group comparison of ANGPTL3 adjusted to age, and sex: Shown are the levels of Total cholesterol (A), HDL-C (B), TG (C), VLDL-p (D), HDL-p (E), and LDL-p (F) in non-carriers, heterozygous and homozygous carrier of ANGPTL3 LOF. P-values from GEE Wald test statistics for pairwise mutation status comparisons are reported, adjusted for multiple comparisons with the Holm method *P<0.05, **P<0.01, ***P<0.001,
-c=cholesterol, -p=particle, Conc.= Concentration.



Supplementary Figure 3: Scatterplot with univariate GEE linear regression slopes (blue) of ANGPTL3 levels (after excluding ANGPTL3 LOF homozygous) against Total cholesterol (A), HDL-C (B), TG (C), VLDL-p (D), HDL-p (E), and LDL-p (F). r = Pearson's correlation coefficients, b=univariate GEE regression coefficients. p-values are from Wald test for the slope. -c=cholesterol, -p=particle, Conc.= Concentration



Supplementary Figure 4: Scatterplot with nonparametric smoothed Local Polynomial Regression fit (green) and Multivariate Adaptive Regression Splines model fit (red) of ANGPTL3 levels (after excluding ANGPTL3 LOF homozygos) against: Total cholesterol (A), HDL-c (B), TG (C), HDL-p (D), VLDL-p (E), and LDL-p (F). -c=cholesterol, -p=particle, Conc.= Concentration



Supplementary figure 5. (A) Linear correlation between LDL-c and PCSK9 levels. (B) Linear correlation between apoB and PCSK9 levels. (C) Linear correlation between LDL-c and apoB levels. (D) Linear correlation between apoB and LDL-bound PCSK9 levels. r = Pearson's correlation coefficients.

	Non-carriers vs. Heterozygotes	Non-carriers vs. Homozygotes	Heterozygotes vs. homozygotes
Total cholesterol (mg/dl)	0.0018	<.001	<.001
HDL-c (mg/dl)	<.001	<.001	<.001
TG (mg/dl)	0.92	<.001	<.001
VLDL-p conc. (nmol/L)	<.001	<.001	<.001
HDL-p conc. (nmol/L)	0.029	<.001	<.001
LDL-p conc. (nmol/L)	0.097	<.001	<.001
PCSK9 (ng/ml)	0.025	<.001	0.0069
LDL-bound PCSK9 (RLU)	0.95	0.0013	0.0059

Supplementary Table 1: Summary table of P-values from Wald test statistics for pairwise mutation status comparisons (group comparison for ANGPTL3 cohort) using generalized estimate equation (GEE) model. -c=cholesterol, -p=particle, conc.=concentration, RLU=Relative Light Units.

	Pearson's correlation coefficients	Regression coefficient	P-value
Total cholesterol (mg/dl)	0.37	0.076	0.0119
HDL-c (mg/dl)	0.33	0.037	0.00465
TG (mg/dl)	0.22	0.036	0.269
VLDL-p conc. (nmol/L)	0.42	0.047	<.001
HDL-p conc. (nmol/L)	0.26	0.01	<.001
LDL-p conc. (nmol/L)	0.3	0.288	0.0277

Supplementary Table 2: Table of Pearson's correlation coefficients, GEE regression coefficients, adjusted for age, denoting the blue slopes on the scatterplots in Figure 1. P-values are from the Wald test for the slopes. -c=cholesterol, -p=particle, conc.=concentration.

	MARS regression r^2	MARS regression equation
Total cholesterol (mg/dl)	0.49	189 + (-1.27)*max(0,60.4-ANGPTL3.ng.dl)+(-2.62)*max(0,40-age)
HDL-c (mg/dl)	0.30	72 + (-0.437)*max(0,60.4-ANGPTL3.ng.dl)+(-0.486)*max(0,age-34)+(-1.48)*max(0,34-age) +(1.29)*max(0,age-66)
TG (mg/dl)	0.14	95.1 + (-0.615)*max(0,81-age)+(-0.508)*max(0,60.4-ANGPTL3.ng.dl)
VLDL-p conc. (nmol/L)	0.29	75.3 + (0.0368)*max(0,ANGPTL3.ng.dl-137.7)+(-0.131)*max(0,137.7-ANGPTL3.ng.dl)+(-0.815)*max(0,76-age)+(-0.851)*max(0,age-38)
HDL-p conc. (nmol/L)	0.21	30.9 + (-0.134)*max(0,61.4-ANGPTL3.ng.dl)+(-0.272)*max(0,72-age)+(-0.23)*max(0,age-34)
LDL-p conc. (nmol/L)	0.25	865 + (-16.9)*max(0,38-age)+(-3.06)*max(0,61.4-ANGPTL3.ng.dl) +(0.569)*max(0,ANGPTL3.ng.dl-266.3)

Supplementary Table 3: Table of MARS model fit R-squared (r^2) and and equations of regression lines denoting the red slopes on the scatterplots in Figures 3 and 4. The equations represent a piecewise function of lines with slopes as the coefficients as above in the form: (slope)*(line equation for hinge function between change-points/knots). -c=cholesterol,-p=particle, conc.=concentration.

	Controls	FHBL1	P-value
Total cholesterol (mg/dl)	204 (56)	102 (18)	<.001
HDL-c (mg/dl)	55 (13)	52 (14)	0.789
TG (mg/dl)	110 (59)	54 (63)	0.0396
LDL-c (mg/dl)	128 (49)	40 (20)	<.001
Apolipoprotein B (mg/dl)	93 (28)	30 (18)	<.001
Total PCSK9 (ng/ml)	206 (35)	173 (45)	0.0165
ANGPTL3 (ng/dl)	235 (110)	211 (77)	0.451

Supplementary Table 4. Characteristics of FHBL type 1 cohort**.** Data presented as means with standard deviation in parenthesis. N=40 subjects (29 FHBL1 with characterized apoB truncation and 11 family-related, normolipidemic controls). -c=cholesterol.

Supplemental Methods - Statistical Analysis Descriptions

Multiple GEE linear regression analyses: Multiple regression models were fit with generalized estimating equations (GEE) to estimate the association for plasma measurements to ANGPTL3 levels and mutation status, adjusting for age, sex, and the correlation within families (1, 2, 3). Analyses were preformed for the full ANGPTL3 cohorts (127 samples) and without the ANGPTL3 LOF homozygous (120 samples). The GEE method estimates standard errors of regression parameters allowing for correlation within families. We would expect that members of the same family would have similar genetic traits as well as environmental factors and hence their plasma measurement values would be correlated more than two unrelated individuals. A working independent covariance structure for family was used. The analysis was performed with the R package “geepack” and regression parameters were tested with a Wald Test (1, 2, 3). All regression estimates for ANGPTL3 were adjusted for age and sex.

Multivariate Adaptive Regression Splines analyses: To study the effect of low vs. high ANGPTL3 levels on plasma measurements, we used a Multivariate Adaptive Regression Spline (MARS) model that allows for different coefficients (or slopes) of ANGPTL3 for low and high values to represent a “threshold” effect. MARS is a non-parametric extension of regression that, through a series of model selection steps, estimates a piecewise linear function that is made up of possibly varying change points or “knots” and slopes that represent a more complex association between ANGPTL3 and the measurements of interest. MARS builds a model that is a linear combination of hinge functions that represent discrete line segments (4,5). This will allow us to interpret the possibly varying associations for different subsets of ANGPTL3 levels. If ANGPTL3 is not significantly associated with the measurements at the specified level of significance of type I error controlled at 5%, it will not select ANGPTL3 into the model. It also selects and adjusts for age if age is a significant confounder. The number of knots in the model was restricted to 8 to reduce over-fitting. Models were implemented using the R package “earth.” The MARS model begins the model building procedure with a set of basis or hinge functions for each data point of the predictor ANGPTL3 (4,5). The procedure then selects the basis functions that fit the data most accurately. For many plasma measurements, the same basis function and hence knot point was selected since the best fitting knot was in approximately the same place. In the scatterplots, the regression lines from the MARS models are visually compared to a GEE model with ANGPTL3, age, and family, as well as a nonparametric smoothed loess curve fit to the ANPTL3 data (4,5). Analyses were preformed for the full ANGPTL3 cohorts (127 samples) and without the ANGPTL3 LOF homozygous (120 samples)

References

1. Højsgaard, S., Halekoh, U. & Yan J. (2006) The R Package geepack for Generalized Estimating Equations Journal of Statistical Software, 15, 2, pp1–11

2. Yan, J. & Fine, J.P. (2004) Estimating Equations for Association Structures Statistics in Medicine, 23, pp859–880.

3. Yan, J (2002) geepack: Yet Another Package for Generalized Estimating Equations R-News, 2/3, pp12-14.

4. Friedman, J. H. (1991). “Multivariate Adaptive Regression Splines”. The Annals of Statistics. 19: 1. doi:10.1214/aos/1176347963

5. Stephen Milborrow. Derived from mda:mars by Trevor Hastie and Rob Tibshirani. Uses Alan Miller’s Fortran utilities with Thomas Lumley’s leaps wrapper. (2016). earth: Multivariate Adaptive Regression Splines. R package version 4.4.4. http://CRAN.R-project.org/package=earth