

1 Human lipoproteins comprise at least 12 different classes that are 2 lognormally distributed.

3

4 Tomokazu Konishi ¹, Risako Fujiwara ^{2,3}, Tadaaki Saito ¹, Nozomi Satou ¹, Naoko Crofts ¹, Ikuko Iwasaki ¹,
5 Yoshihisa Abe ², Shinpei Kawata ², and Tatsuya Ishikawa ²

6 ¹ Faculty of Bioresource Sciences, Akita Prefectural University

7 ² Research Institute of Akita Cerebrospinal and Cardiovascular Center, Akita Prefectural Hospital
8 Organization

9 ³ Cardiovascular internal medicine, Akita City Hospital

10

11 Summary

12 Lipoproteins in medical samples have been measured by enzymatic methods that coincide with conventional
13 ultracentrifugation. However, the high gravity and time required for ultracentrifugation can cause sample
14 degradation. This study presents the results of HPLC, a gentler and rapid separation method, for 55 human
15 serum samples. The elution patterns were analysed parametrically, and the attribute of each class was
16 confirmed biochemically. Human samples contained 12 classes of lipoproteins, each of which may consist
17 primarily of proteins. There are three classes of VLDLs. The level of each class was distributed lognormally,
18 and the standard amount and the 95% range were estimated. Enzymatic methods measure the levels of several
19 mixed classes. This lognormal character suggests that the levels are controlled by the synergy of multiple
20 factors.

21

22

23 **Introduction**

24 Lipoproteins are measured in two ways. The first is related to class separation for biochemical purposes ¹⁻⁴.
25 This method uses ultracentrifugation, which spontaneously creates a salt density gradient due to centrifugal
26 forces. Although the classes of lipoproteins are separated according to their density, this method is time-
27 consuming and is not suitable for measuring large numbers of medical samples. The other method, which uses
28 enzymatic processes, is used for physical examination ⁵. In this method, cholesterol is chemically extracted
29 from a certain class of lipoproteins and then measured. Several kits are available, but all are adjusted to mimic
30 the results of the ultracentrifugation method. In the ultracentrifugation approach, the larger the particle size,
31 the lower the density of lipoproteins; therefore, names such as very low-density lipoprotein (VLDL) and high-
32 density lipoprotein (HDL) are used for large and small particles, respectively ^{6,7}.

33
34 The accuracy of ultracentrifugation has been questioned in a study using rat serum ⁸. The very high centrifugal
35 force was sufficient to pull hydrophobic proteins out of the membrane ⁹; additionally, complete separation can
36 take several days, during which proteins can be degraded. There is a gentler way to separate these classes via
37 HPLC gel filtration. This takes up to 30 min and does not require extra salts ¹⁰. This is not a novel method, but
38 there were a few issues with how the data were analysed, as it was noted that the data were intended to be
39 consistent with standard ultracentrifugation results. Analysis of HPLC results of rat data with parametric
40 analysis (Materials and Methods) showed striking differences from the ultracentrifugation results ⁸. All classes
41 of lipoproteins are protein-rich particles, contrary to conventional knowledge ^{6,7}. Two new classes, LDL-
42 antiprotease complexes (LAC), were also discovered.

45 **Results and discussion**

46 At least 12 classes of lipoproteins in normal distribution had to be presumed to fit the human data (Fig. 1);
47 even a minor class cannot be ignored (Fig. S2). The attributes of the classes were determined based on the
48 elution pattern of the major protein components (Fig. 1, S3-S6). The elution patterns of these proteins

49 coincided with the distribution of the corresponding classes. As the positions and scales coincided between
50 TG and cholesterol, it was confirmed that a class had a fixed ratio of these substances.

51

52 Many features in human samples were similar to rat results ⁸. Each elution pattern was a mixture of normal
53 distributions (Fig. 1 and S2-S6). This shows that each class of lipoprotein is stable in serum, which is
54 inconsistent with the scenario in which CM and VLDL lose TG by gradual degradation ^{6,7}, which in turn
55 produces more skewed distributions ⁸. Rather, a single degradation of TG should have converted them into the
56 next class at once.

57

58 CM and VLDL were eluted, consistent with observed peaks of ApoB48 and ApoB100, respectively (Fig. 1A
59 and S3). Most ApoB proteins appear to be degraded during TG removal. In fact, the LDL1 fraction contains
60 less B100 than expected for a large number of particles, and the LDL2 fraction is devoid of B48. HDLs were
61 eluted with ApoA-1 (Fig. S4). Fractions of LAC1 and LAC2 eluted with major antiproteases, Alpha-2-
62 macroglobulin and inter-alpha-trypsin inhibitor, respectively, similar to those in rats (Fig. S5) ⁸. Antiproteases,
63 and thus LACs, are synthesised in the liver ^{11,12}; they may supply cholesterol to the thrombus and protect it
64 from lysis by plasmin.

65

66 There were some differences observed between rat and human lipoproteins. There were probably three classes
67 of VLDL particles in human lipoproteins, which contained ApoB100 (Figs. 1 and S3), compared to rat
68 lipoprotein which only has one class of VLDL particles. These lipoproteins may be secreted by the liver. Here,
69 the largest class is denoted as VLDL. Another cholesterol-rich class included another lipoprotein denoted as
70 Lp(a) protein ^{13,14} (Fig. 2A). The size of the particles seems to be larger than that previously reported ¹³;
71 ultracentrifugation may have removed a portion of the particles. Another smaller class was noted to have a
72 size slightly larger than that of LDL1. Estimating the cholesterol content of this class was difficult as its size
73 is close to that of LDL1. However, as the fitting results suggested low cholesterol (S2 Table), this is denoted
74 as the TG-rich class (TR) in this study. The levels of VLDL, Lp(a), and TR showed no relationship, suggesting

75 that they were produced independently (Fig. S8AB). Rather, the cholesterol carrier, Lp(a), showed a weak
76 negative correlation with the levels of another carrier, LDL2 (Fig. S8C).

77

78 There were two sizes of larger particles with the largest particles with B48 denoted as CM1. The 50-60 nm
79 particles had B48 (Figs. 1, S3, and S6), and the level showed a higher correlation with CM1 than VLDL (Fig.
80 S8DE); hence, the class is denoted as CM2. These particles would be intestine-derived ^{6,7}.

81

82 The particle size and ApoB protein distribution suggest that they can be classified into three lineages:

83 1. CM → LDL2 → LAC2

84 2. VLDL, Lp(a) → LDL1 → LAC1

85 3. HDL2 → HDL1 → mHDL.

86 These protein distributions were the same as those of the rats ⁸. In this study, we focused on determining the
87 attributes; hence, we did not analyse other apoproteins such as ApoC or ApoE. Certainly, we may only identify
88 a fraction of apoproteins ⁸. Additionally, it should also be noted that the number of classes presented here was
89 the smallest to perform curve fitting (Fig. S2). Systems with better separation may find more classes. The
90 positions and scales of the classes are listed in Table S1. Classes that were initially secreted from the liver or
91 intestine showed a larger scale. The particles may have a larger tolerance for the ratio and amount of lipids,
92 which may have expanded this parameter. However, the elution position of the class did not change
93 significantly between the samples. Therefore, the estimated peak diameters were within a certain range,
94 indicating the physical stability of the particles (Table S1).

95

96 The amount of each class varied significantly among the volunteer samples (Fig. 2BC, C). They were
97 lognormally distributed, as was found from the logarithms of the data quantiles, which showed a linear
98 relationship with the theoretical values of the normal distribution. Linear correlations were confirmed in all
99 classes (Fig. S7); The slope and intercept of the regression line represent σ and μ of the normal distribution,
100 respectively; these are the distributions used to fit the curves of each sample. The parameters were estimated

101 by robust methods, specifically MAD and trimmed means; the appropriateness of the estimation can be
102 checked to determine whether the lines fit the plots. By accumulating z-normalized data using these parameters,
103 $z_i = (\log(x_i) - \mu) / \sigma$ for any raw data x_i , the distribution can be confirmed in a more exact quantile-quantile (QQ)
104 plot (Fig. 3), where the slope is 1 and the intercept is zero. The lower part of the QQ plot bends downward,
105 which is considered to be an artefact during curve fitting because low values are easy to ignore.

106

107 It is not surprising that each class was lognormally distributed. For example, the transcriptome has the same
108 distribution properties ¹⁵. Levels of lipoproteins, as well as mRNA, are regulated by a balance between
109 synthesis and degradation, both of which are controlled by specific factors in a multiplicative manner. This
110 mechanism determines the distribution (Fig. S11). Therefore, lipoprotein levels can easily change due to
111 multiple small causes; differences in multiple lifestyle habits may worsen the medical condition. In contrast,
112 simultaneous clinical efforts may synergistically improve lipoprotein levels. This would explain why a
113 different set of feeds drastically altered lipoprotein levels in a rat study ¹⁶.

114

115 Knowing that the data is lognormally distributed, the standard value and the 95% range of the classes could
116 be estimated with accuracy, even though they were largely dispersed (Table S2). Note that the interval ranges
117 inevitably become asymmetric to the standards: the higher is always wider (Fig. 11B).

118

119 The ratio of TG to cholesterol in each class also fluctuated greatly (Fig. 2A). Naturally, these ratios are large
120 in the class synthesised to include TG, such as CM and VLDL. HDL1 and 2 are considered precursors to HDL
121 (Fig. S4). They mature by receiving cholesterol.

122

123 LDL1 was the most predominant cholesterol carrier, followed by LAC2 (Fig. 2B). Some cholesterol carriers
124 appeared within certain narrow ranges (LDL1, LACs, and HDLs). These classes may maintain cholesterol
125 homeostasis. Conversely, variations in CM and VLDL levels were particularly high (Fig. 2B). The former
126 may depend on diet, and the latter may depend on the body's requirements for TG as a storable energy source.

127 Among the TG-rich classes, TR appeared to be fairly stable.

128

129 HDLs were minor cholesterol carriers, contradicting the estimations of the previous study, which did not check
130 whether each of the fractions contained ApoA-1 protein ¹⁷. In the study by Gordon et al., they assigned the
131 whole of a major peak of cholesterol as HDL, which seemed to be a reasonable decision, since HDL was
132 believed to be the major cholesterol carrier from ultracentrifugation results ^{6,7}. However, in reality, the peak
133 size is too large to be assigned to HDL alone; the structure of HDL is surrounded by ApoA-1 and therefore
134 has a limitation in size ¹⁸. In fact, ApoA-1 was only confirmed at the end of the peak (Fig. S4). Rather, the
135 peak was mainly composed of LDL2 and LAC2 (Fig. 1); these classes may have behaved as high-density
136 particles during ultracentrifugation. However, their origins and functions are quite different from those of HDL.

137

138 In contrast to rats, human serum contained higher amounts of immunoglobulins. It should be noted that rats
139 grown in pathogen-free environments did not show detectable levels of these proteins in SDS-PAGE ⁸. Some
140 of them were as large as certain classes of lipoproteins (Fig. S9), which interfered with the estimation of
141 protein abundance in these classes.

142

143 In contrast, CM and VLDL are very large particles; hence, the corresponding fractions would be free from
144 immunoglobulins. The amount of protein was estimated from the UV record of preparative HPLC, and the
145 ratios of the lipids (TG and cholesterol) were observed (Fig. 2D). The contents of both classes varied greatly;
146 however, they always contained fewer lipids than protein, contradicting the results of ultracentrifugation ^{6,7}.

147

148 The amounts of HDL and LDL measured by conventional enzymatic methods were much higher than those
149 of any of the classes. This is not surprising, as those account for the majority of total cholesterol (Fig. S10). If
150 conventional methods extract cholesterol from certain classes with high efficiency, the observed data would
151 reflect combinations of some classes. Such combinations were estimated as combinations with the highest
152 correlation and similarity (Fig. 4). The conventional measure of LDL would be VLDL derivatives without

153 VLDL, and HDL would be HDLs and CM derivatives without CM1. Of course, we cannot deduce any of the
154 true classes from the conventional measurements. In particular, the estimated HDL differed significantly from
155 the actual amount of HDLs. Moreover, sums of lognormally distributed quantities are not informative at all,
156 as an anomalous class may make up the majority of the sum; in a statistical sense, they do not even show
157 average levels (Fig. S12). Unfortunately, it is also true that total TG or cholesterol does not provide useful
158 information. Rather, an exact measurement of each class is desirable.

159

160 Thus, the perception of lipoproteins should be updated. They are classes of protein-rich particles, each of
161 which has specific functions. Only limited apoproteins were studied here, but the attributes of other
162 apoproteins are important for their functions. Hence, the classes need to be studied extensively, which will
163 provide a deeper understanding of the pathophysiology. The levels of the classes varied among volunteers and
164 were lognormally distributed. Some classes are narrowly controlled and are good candidates for indicators of
165 diseases. Conventional enzymatic methods measure mixtures of multiple classes. Because the sums of
166 lognormally distributed numbers are not informative (Fig. S12), the data do not provide a proper
167 diagnostic criterion. This could be the reason why levels of conventional LDL did not indicate the prognosis
168 of patients¹⁹. However, HPLC is not suitable for large numbers of samples. Therefore, simpler methods for
169 measuring specific classes are required. In particular, the targets of enzymatic methods should be completely
170 refined. According to the lognormal distribution characteristics, independent clinical treatments, such as
171 antilipidemic drugs, nutritional therapy, and ergotherapy, may synergistically change the levels of specific
172 classes.

173

174

175 Acknowledgements

176 We would like to thank Editage (www.editage.com) for English language editing.

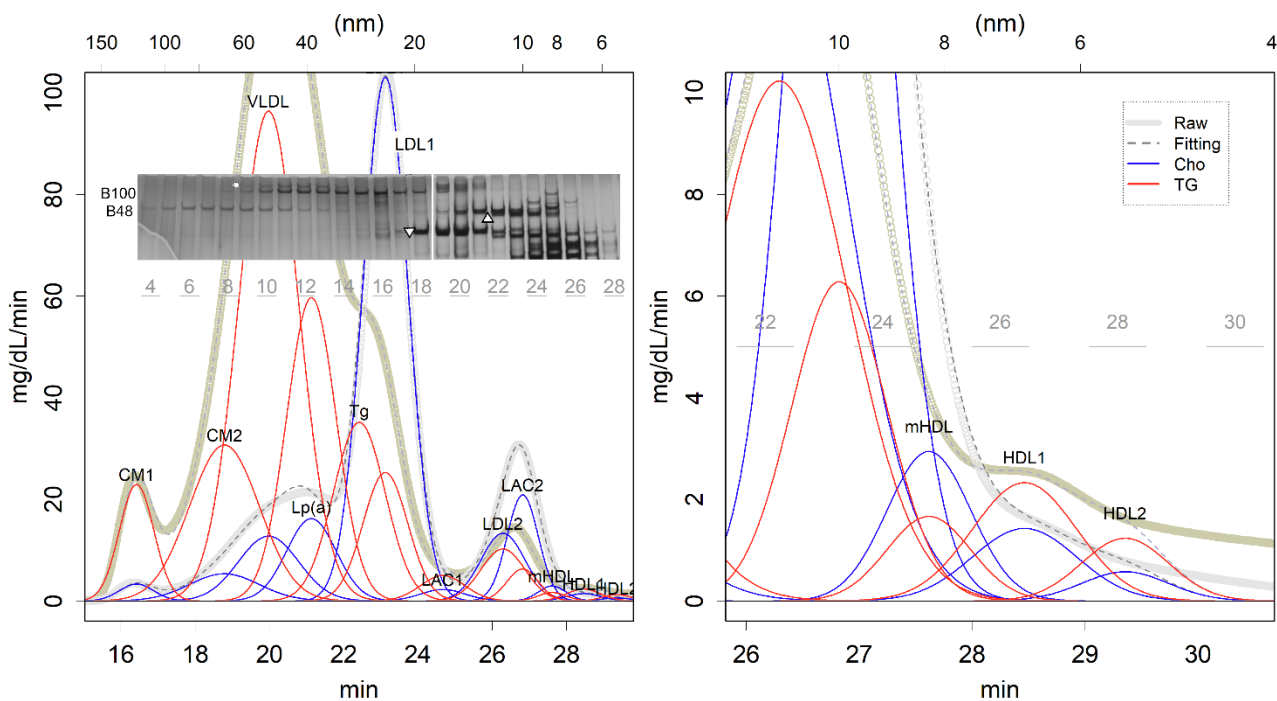
177

178

179 References

- 180 1 Redgrave, T. G., Roberts, D. C. & West, C. E. Separation of plasma lipoproteins by
181 density-gradient ultracentrifugation. *Anal Biochem* **65**, 42-49 (1975).
- 182 2 Chapman, M. J., Goldstein, S., LAGRANGE, D. & LAPLAUD, P. M. A density gradient
183 ultracentrifugal procedure for the isolation of the major lipoprotein classes from human
184 serum. *J Lipid Res* **22**, 339-358 (1981).
- 185 3 Oschry, Y. & Eisenberg, S. Rat plasma lipoproteins: re-evaluation of a lipoprotein system
186 in an animal devoid of cholesteryl ester transfer activity. *J Lipid Res* **23**, 1099-1106
187 (1982).
- 188 4 Stahlman, M. *et al.* Proteomics and lipids of lipoproteins isolated at low salt
189 concentrations in D₂O/sucrose or in KBr. *J Lipid Res* **49**, 481-490,
190 doi:10.1194/jlr.D700025-JLR200 (2008).
- 191 5 Bachorik, P. S. & Ross, J. W. National Cholesterol Education Program recommendations
192 for measurement of low-density lipoprotein cholesterol: executive summary. The
193 National Cholesterol Education Program Working Group on Lipoprotein Measurement.
194 *Clin Chem* **41**, 1414-1420 (1995).
- 195 6 Voet, D. & Voet, J. G. *Biochemistry*. 3 edn, (Wiley, 2004).
- 196 7 Berg, J., Tymoczko, J. L. & Stryer, L. Ch. 742–748, (W. H. Freeman, 2006).
- 197 8 Konishi, T. & Takahashi, Y. Lipoproteins comprise at least 10 different classes in rats,
198 each of which contains a unique set of proteins as the primary component. *PLoS ONE*
199 **13**, e0192955 (2018).
- 200 9 Ma, C. D., Wang, C., Acevedo-Velez, C., Gellman, S. H. & Abbott, N. L. Modulation of
201 hydrophobic interactions by proximally immobilized ions. *Nature* **517**, 347-350,
202 doi:10.1038/nature14018 (2015).
- 203 10 Okazaki, M. & Yamashita, S. Recent advances in analytical methods on lipoprotein
204 subclasses: calculation of particle numbers from lipid levels by gel permeation HPLC
205 using "spherical particle model". *J Oleo Sci* **65**, 265-282, doi:10.5650/jos.ess16020 (2016).
- 206 11 Naveau, S., Poynard, T., Benattar, C., Bedossa, P. & Chaput, J. C. Alpha-2-
207 macroglobulin and hepatic fibrosis. Diagnostic interest. *Digestive diseases and sciences*
208 **39**, 2426-2432, doi:10.1007/bf02087661 (1994).
- 209 12 Kim, T. H. *et al.* Overproduction of inter- α -trypsin inhibitor heavy chain 1 after loss of
210 $G\alpha_{13}$ in liver exacerbates systemic insulin resistance in mice. *Science*
211 *Translational Medicine* **11**, eaan4735, doi:10.1126/scitranslmed.aan4735 (2019).
- 212 13 Berg, K. A new serum type system in man -- the Lp system. *Acta pathologica et*
213 *microbiologica Scandinavica* **59**, 369-382, doi:10.1111/j.1699-0463.1963.tb01808.x (1963).
- 214 14 Maranhão, R. C., Carvalho, P. O., Strunz, C. C. & Pileggi, F. Lipoprotein (a): structure,
215 pathophysiology and clinical implications. *Arq Bras Cardiol* **103**, 76-84,
216 doi:10.5935/abc.20140101 (2014).
- 217 15 Konishi, T. A thermodynamic model of transcriptome formation. *Nucleic acids research*
218 **33**, 6587-6592, doi:10.1093/nar/gki967 (2005).
- 219 16 Takahashi, Y., Konishi, T. & Yamaki, K. Tofu and fish oil independently modulate serum
220 lipid profiles in rats: Analyses of 10 class lipoprotein profiles and the global hepatic
221 transcriptome. *PLoS ONE* **14**, e0210950 (2019).
- 222 17 Gordon, S. M., Deng, J., Lu, L. J. & Davidson, W. S. Proteomic characterization of human
223 plasma high density lipoprotein fractionated by gel filtration chromatography. *J*
224 *Proteome Res* **9**, 5239-5249, doi:10.1021/pr100520x (2010).
- 225 18 Wu, Z. *et al.* The low resolution structure of ApoA1 in spherical high density lipoprotein
226 revealed by small angle neutron scattering. *J Biol Chem* **286**, 12495-12508,
227 doi:10.1074/jbc.M110.209130 (2011).

- 228 19 Goff, D. C., Jr. *et al.* 2013 ACC/AHA guideline on the assessment of cardiovascular risk:
229 a report of the American College of Cardiology/American Heart Association Task Force
230 on Practice Guidelines. *J Am Coll Cardiol* **63**, 2935-2959, doi:10.1016/j.jacc.2013.11.005
231 (2014).
- 232 20 Tomokazu, K. *Human lipoproteins comprise at least 12 different classes that are*
233 *lognormally distributed*, <<https://doi.org/10.6084/m9.figshare.14247119.v1>> (2021).
- 234 21 Ellis, G. & Silk, J. Scientific method: Defend the integrity of physics. *Nature* **516**, 321-
235 323, doi:10.1038/516321a (2014).
- 236
- 237



238

239

240

241

242

243

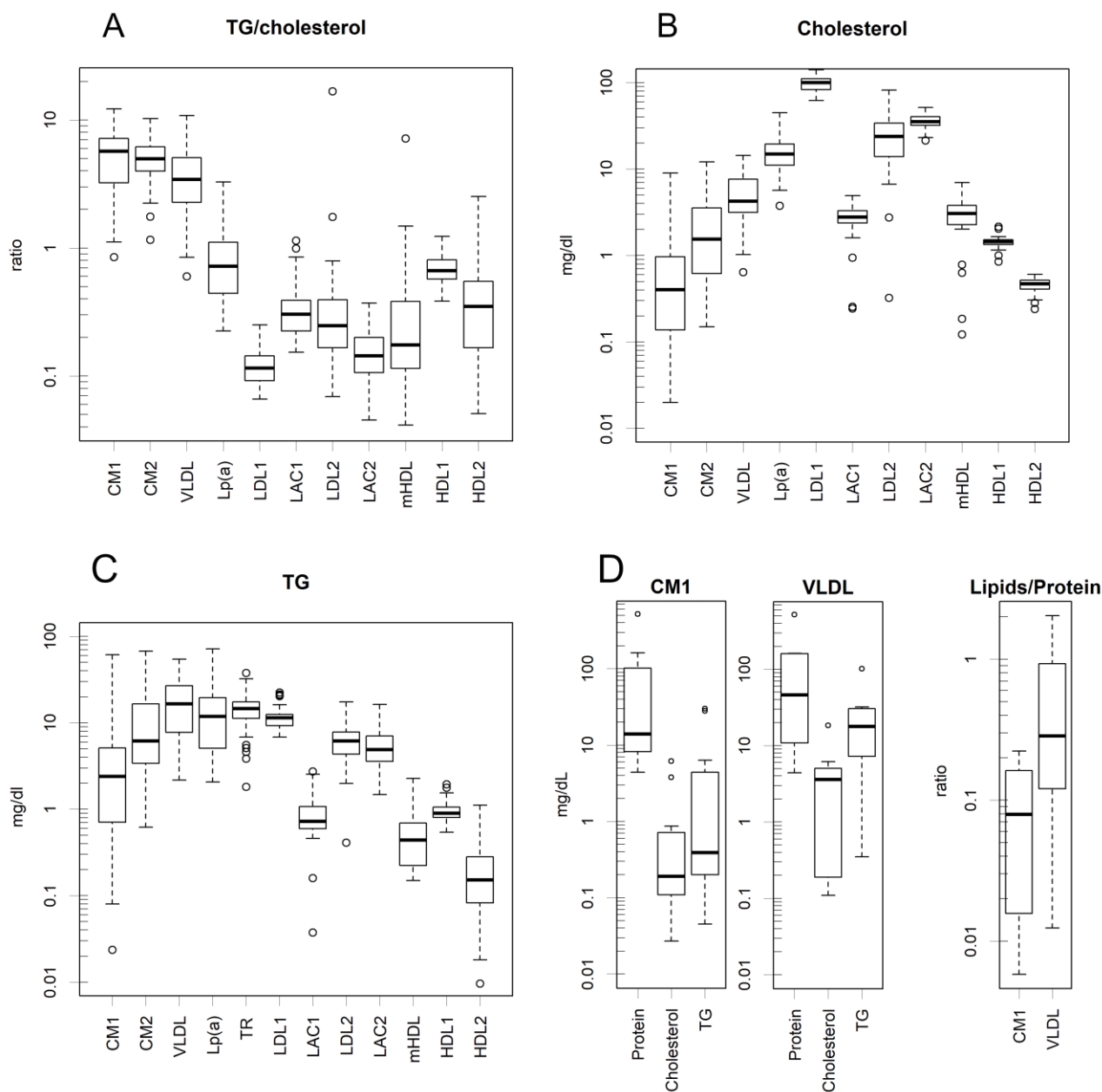
244

245

246

247

Fig. 1 Elution pattern of HPLC of a hyperlipidaemia patient. In gel filtration, the shorter the elution time, the larger the particle size; elution time is proportional to the logarithms of the diameter of the particles. Left: whole image, right: enlarged view around HDLs. The bold line presents the measurement raw data, the red (TG) and blue (cholesterol) lines are the curve-fitted classes, and the dotted lines are their sum. There are 12 classes. The superimposed photo is a part of SDS-PAGE. B100 and B48 are the respective ApoB positions. ▽: Alpha-2-macroglobulin, △: Inter-alpha-trypsin inhibitor heavy chain, ○: Lp(a). Here an example of a hyperlipidaemic patient is shown so that the classes can be easily observed. An alternative to healthy volunteers is shown in S6 Fig.

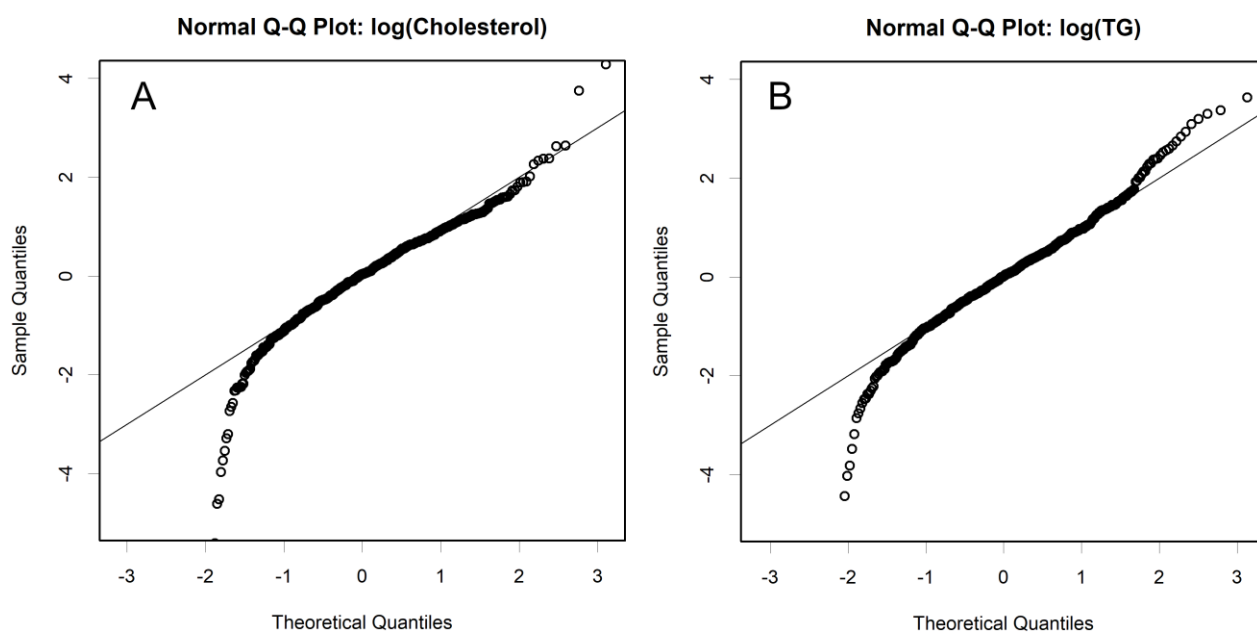


248

249 **Fig. 2 Box plot of the logarithm of quantity.** All the y-axes are on a logarithmic scale. **A.** The ratio of
 250 TG/cholesterol. Naturally, the ones produced with TG are higher. The contents are **B.** Cholesterol and **C.** TG
 251 (mg / dL). While LDL1, LACs, and HDLs appeared within a certain narrow range, CM and VLDL levels
 252 fluctuated. **D.** Amounts and ratios of CM and VLDL measured by the preparative HPLC of healthy samples.
 253 Lipids (TG + cholesterol) are less than proteins in those classes.

254

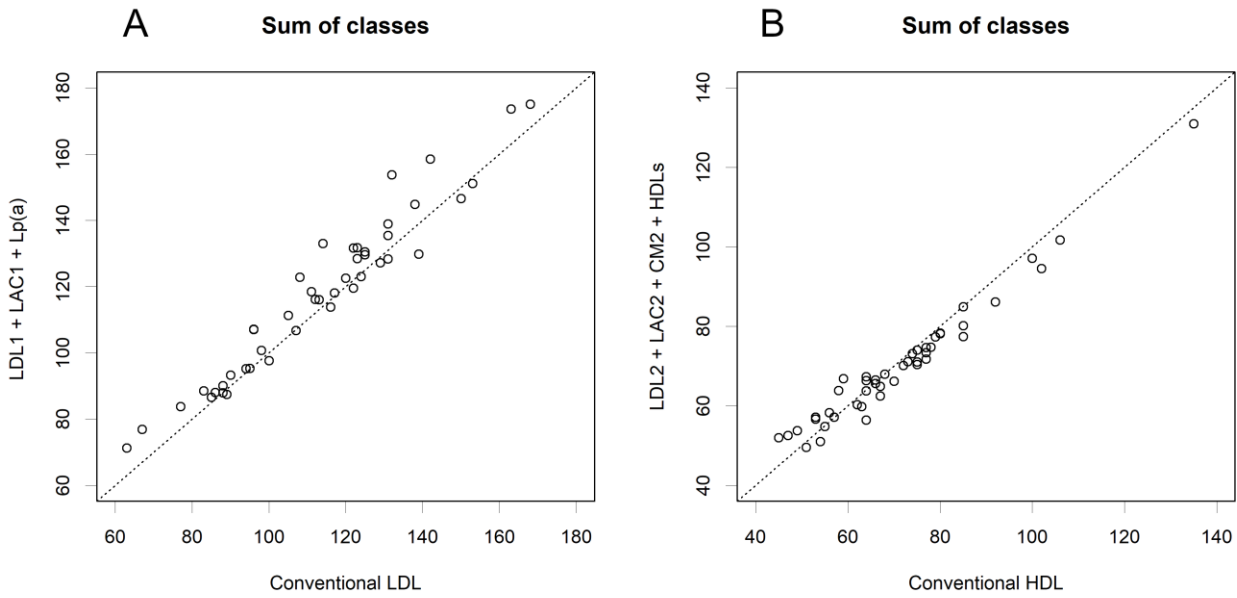
255



256

257 **Fig. 3 Normal Q-Q plot of the logarithms of normalized data. Cholesterol (A.) and TG (B).** Quantiles are
258 compared between the theoretical normal distribution and the z-normalized log-data. Here a lognormal
259 distribution will form a straight relationship.

260



261

262

263

264

265

266

Fig. 4 Estimation of the sets of classes that are measured by a conventional method. If the conventional methods extract most of the cholesterol from particular classes, these would present the sum of the quantities of the classes. The combinations presented here were the ones with the highest correlation and the closest amounts.

267 Supplementary Materials

268 Supplementally figures and tables are in Figshare ²⁰.

269

270 **Materials and Methods**

271 Blood samples: Samples were collected after obtaining informed consent from all volunteers and approval
272 from the ethics committee of Akita Cerebrospinal and Cardiovascular Center (ID. 19-21). All samples were
273 anonymized prior to analysis. No postmeal time was specified for blood collection. The age of the volunteers
274 is shown in Figure S1.

275

276 The collected blood was sent to a clinical laboratory (SRL Inc., Tokyo, Japan). The serum was separated, and
277 HLD, LDL, and total cholesterol were measured using conventional enzymatic methods. The serum was sent
278 to Skylight Biotech Inc. (Akita, Japan) for further analysis using gel filtration HPLC. Forty-four healthy
279 volunteer samples were subjected to analytical HPLC, and TG and cholesterol were monitored sequentially.
280 Furthermore, 11 samples (6 healthy and 5 hyperlipidemic) were subjected to preparative HPLC and then
281 fractionated.

282

283 The HPLC monitoring data (TG and cholesterol) were analysed parametrically ⁸. This method is a
284 parsimonious way of performing curve fit to maintain the falsifiability of the model using the minimum
285 number of classes assumed ²¹. With many estimated classes, the fitting process will become easier;
286 however, the assumed classes must be verified by reality. Too many assumptions make this verification
287 difficult. The size and range of a class are presented using the position μ and scale σ of the normal distribution.
288 We assumed that a class would contain TG and cholesterol at a certain constant rate regardless of the size
289 differences within the class., with The amount of each TG and cholesterol presented by using another
290 parameter. This assumption was verified through the curve-fitting process.

291

292 The standard values of TG and cholesterol in each class were estimated from the full data set using a trimmed
293 mean (0.2). Their 95% range was estimated using the median absolute deviation (MAD): the upper and lower
294 limits of the 44 healthy samples were estimated as trimmed mean to two MADs. The standard values of the
295 position or scale parameters of the classes, which were varied, were estimated from the trimmed mean of μ or
296 σ^2 found in each sample.

297

298 In preparative HPLC, the elution was periodically fractionated. UV was also monitored, and the protein
299 amounts of CM and VLDL were estimated under the assumption that any protein absorbs a fixed amount of
300 UV per weight. Each fraction was subjected to 5 %–20% SDS-PAGE, and the proteins were detected using
301 silver staining. Some protein bands were identified using MALDI-TOF MS (Genomine Inc., Kyungbuk,
302 Korea). In addition, specific proteins were confirmed by western blotting after transfer to PVDF membranes.
303 The antibodies used were as follows: anti-apoB antibody (A-6), sc-393636 AF488; apoA-I antibody (B-10),
304 sc-376818 AF647 (Santa Cruz Biotechnology Inc., Texas, USA); anti-Lipoprotein a antibody, ab27631,
305 (Abcam plc., London, UK). Chemiluminescence of the antibodies and silver-stained gel bands were measured
306 using an Amersham Typhoon Scanner (Cytiva).

307