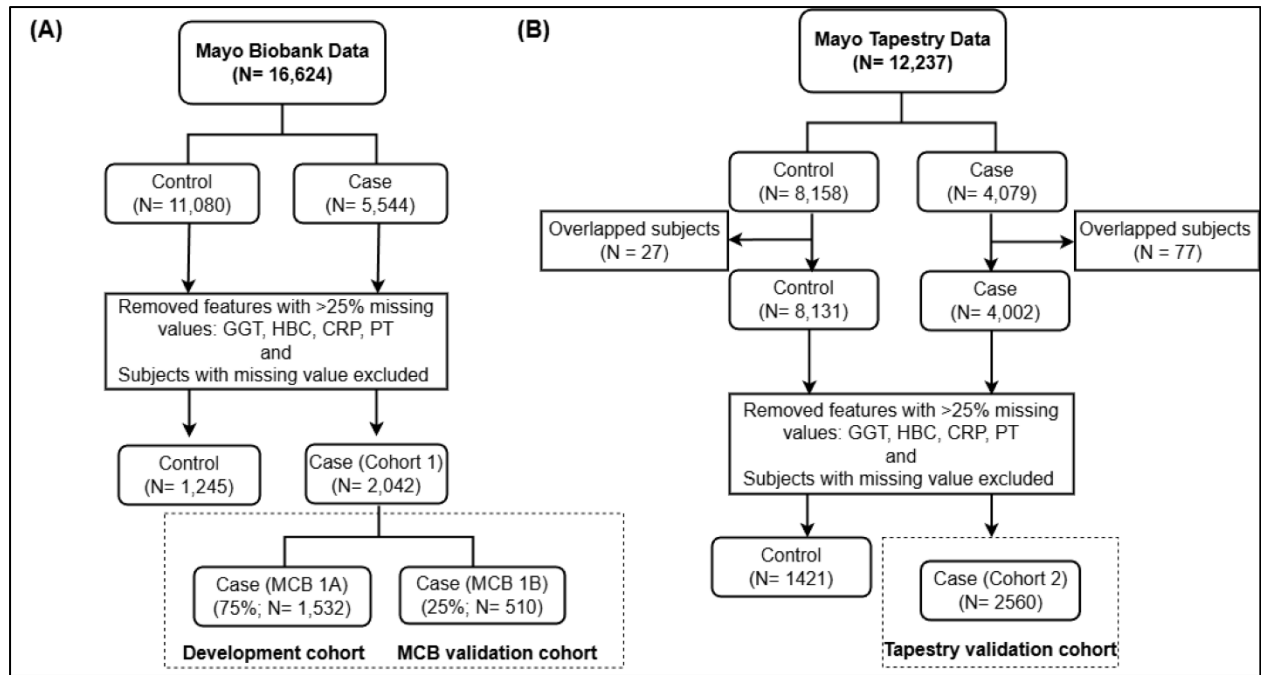


Subtyping metabolic dysfunction-associated steatotic liver disease using electronic health record-linked genomic cohorts reveals diverse etiologies and progression

Supplementary information	Description
Supplementary Fig. 1	Workflow for constructing MASLD case and control cohorts
Supplementary Fig. 2	Bivariate residual (BVR) matrix for the final five-class LCA model in the MCB development cohort
Supplementary Fig. 3	Distributions of PRSs for MASLD and four comorbidities
Supplementary Fig. 4	Liver stiffness (LS) levels varied by subtype in the two cohorts
Supplementary Fig. 5	Reproducibility of probability- and core points-based subtype assignments in the MCB validation cohort
Supplementary Fig. 6	Reproducibility of probability- and core points-based subtype assignments in the Tapestry validation cohort
Supplementary Fig. 7	LCA-identified subtypes vs. subtypes assigned in the two validation cohorts
Supplementary Fig. 8	The cumulative risks of developing intrahepatic and extrahepatic complications stratified by subtype based on probability-based assignment in Tapestry cohort
Supplementary Fig. 9	The cumulative risks of developing intrahepatic and extrahepatic complications stratified by subtype from centroid-based assignment in Tapestry cohort
Supplementary Fig. 10	The cumulative risks of developing intrahepatic and extrahepatic complications stratified by subtype from core points-based assignment in Tapestry cohort
Supplementary Fig. 11	Changes of clinical biomarkers over 6-year follow-up in Tapestry cohort
Supplementary Fig. 12	Medication across subtypes in Tapestry cohort
Supplementary Fig. 13	Association of subtypes with medication stratified by liver disease and comorbidities in Tapestry cohort
Supplementary Fig. 14	The changes of four clinical biomarkers upon medication in Tapestry cohort
Supplementary Fig. 15	Proportion of subjects with top decile of PRSs in the assigned subtypes and control in Tapestry cohort
Supplementary Fig. 16	Enrichment of key MASLD genetic variants across subtypes and control
Supplementary Fig. 17	MASLD partitioned PRSs and the associations with liver and cardiometabolic diseases in the MCB cohort
Supplementary Fig. 18	Cross-study comparison of two subtyping approaches in Tapestry
Supplementary Fig. 19	Heatmaps of similarity matrix between two sets of subgroups in MCB cohort (A) and Tapestry cohort (B)
Supplementary Fig. 20	Consistency of LCA-derived subgroups across development and validation cohorts
Supplementary Table 1	Presence of cardiometabolic risk factors in MASLD patients
Supplementary Table 2	Criteria used for continuous to categorical variable conversion
Supplementary Table 3	LCA model-fit indices in the MCB development cohort

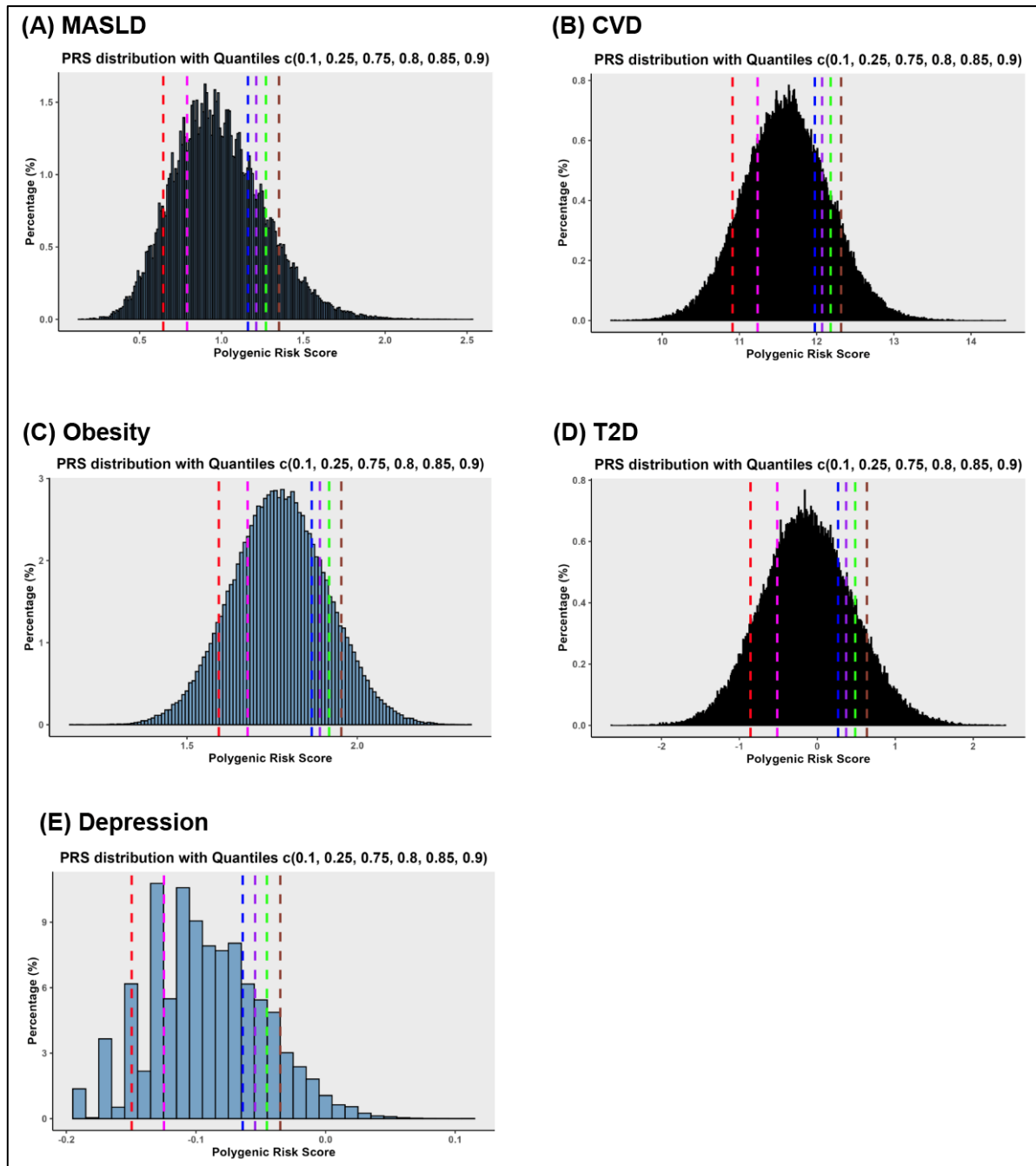
Supplementary Table 4	Baseline characteristics of subgroups in MCB development cohort based on data within one year before and after first diagnosis of MASLD
Supplementary Table 5	Performance metrics of three subgroup assignment methods in the two validation cohorts
Supplementary Table 6	LCA-based vs. assignment-based subgroups in the two validation cohorts
Supplementary Table 7	Accuracy of subgroup assignments in subsets of MASLD patients diagnosed with type 2 diabetes or obesity
Supplementary Table 8	Risk of MASLD related intrahepatic and extrahepatic disease complications across subtypes in MCB cohort
Supplementary Table 9	Changes of biomarkers over six years within each subgroup in the MCB cohort
Supplementary Table 10	Risk of liver transplantation in the two cohorts



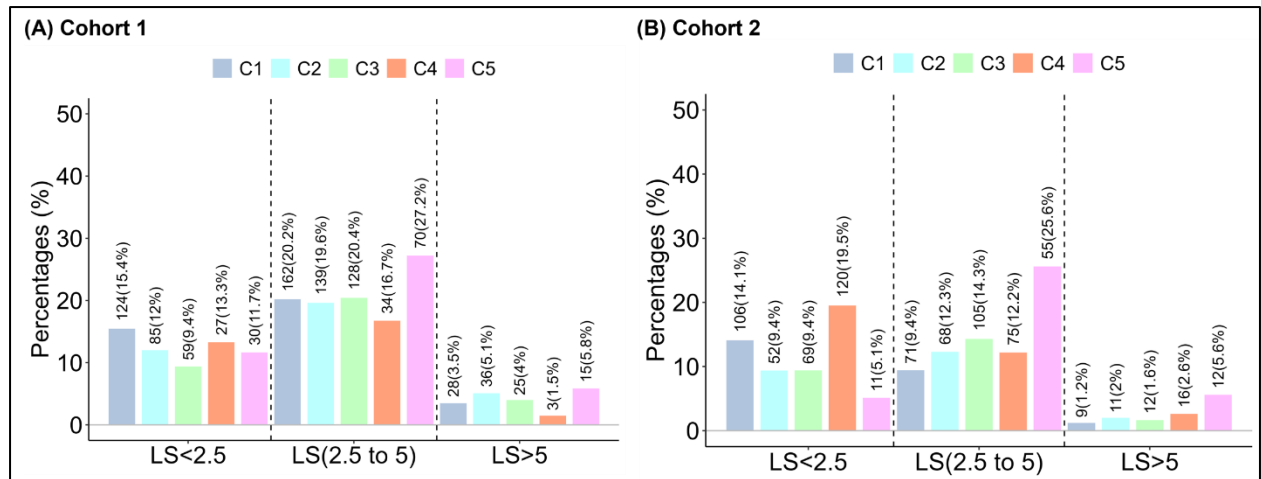
Supplementary Fig. 1. Workflow for constructing MASLD case and control cohorts. a MCB control, development cohort (MCB 1A) and validation cohort (MCB 1B). **b** Tapestry validation cohort. Four features (GGT, HBC, CRP, and PT) were excluded from analysis due to high missingness. For each of these four, data were not available in >25% of the subjects. For Tapestry Study, subjects who were also included in MCB (27 controls and 77 cases) were excluded. In each cohort, subjects with missing value in any of the remaining variables (see Supplementary Data 1) were filtered out. GGT, gamma-glutamyl transferase; HBC, Hepatitis B Core (HBC) total antibody test; CRP, C-reactive protein; PT, prothrombin time; MCB, Mayo Clinic Biobank.

Sex	0.68	0.62	0.02	1.11	5.25	4.74	0.1	0.53	0.3	0.9	15.9	0.29	
Obesity	0.4	0.34	0.01	8.08	1.57	0.67	0.4	1.91	0.01	0.23	0.04		
Migraine	0.91	0.57	0.01	0.27	25.97	0.35	0.76	1.84	0.3	0			
MetS	0.7	0.46	0	0.26	0.42	0.07	6.52	0.06	1				
Kidney disease	18.29	3.72	0.01	2.97	0.66	0.58	0.22	26.63					
Hypertension	1.62	1.63	0.26	0.24	0.03	0.33	3.13						
Hyperlipidemia	0.23	0.05	0.18	0.18	0.84	0.58							
HDL	0.09	0.79	0.25	4.07	0.15								
Depression	1.71	0.06	0.05	1.63									
BMI	0.7	1.07	0.19										
AST	1.28	0.68											
ALT	0.72												
ALP													
	ALP	ALT	AST	BMI	Depression	HDL	Hyperlipidemia	Hypertension	Kidney disease	MetS	Migraine	Obesity	Sex

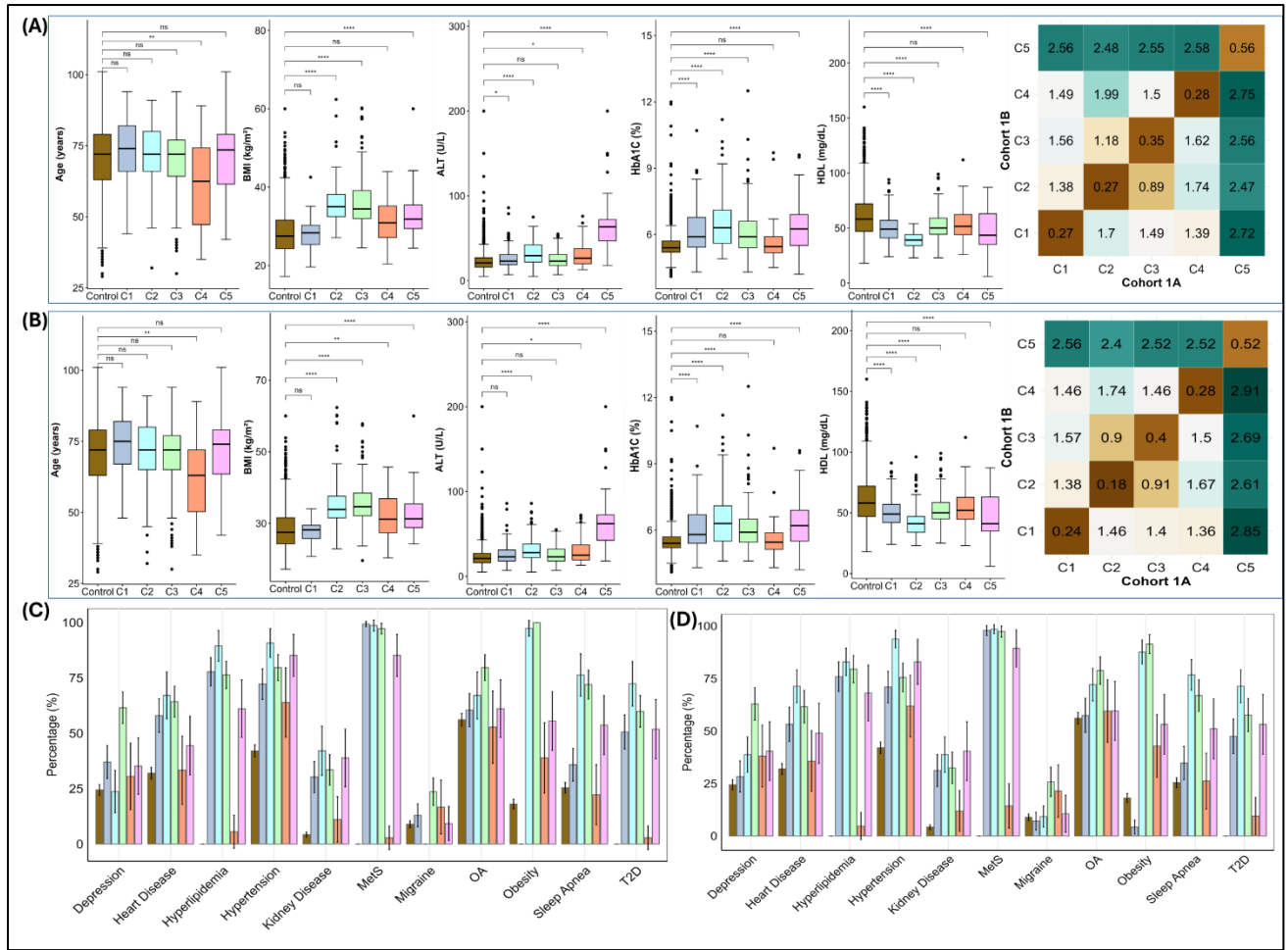
Supplementary Fig. 2. Bivariate residual (BVR) matrix for the final five-class LCA model in the MCB development cohort. The matrix showing BVRs for all 78 pairwise combinations of the 13 indicator variables included in the latent class analysis (LCA) in the development cohort (MCB 1A). These variables include sex, body mass index (BMI), alanine aminotransferase (ALT), aspartate aminotransferase (AST), alkaline phosphatase (ALP), high-density lipoprotein (HDL), obesity, hyperlipidemia, metabolic syndrome (MetS), hypertension, depression, migraine, and kidney disease. Each value represents the BVR for the corresponding indicator pair in the selected five-class solution.



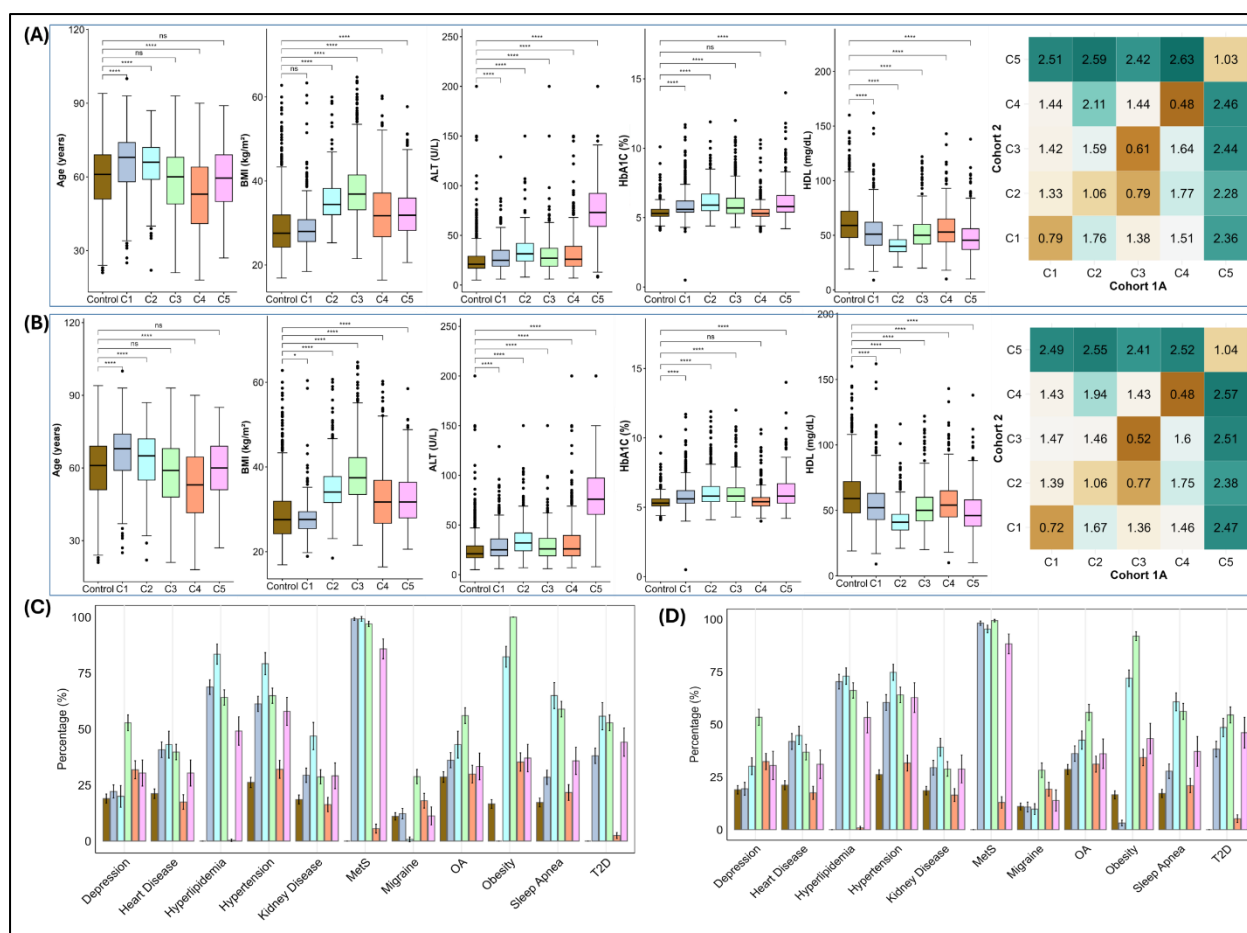
Supplementary Fig. 3. Distributions of PRSs for MASLD and four comorbidities. A MASLD PRS-15. **B** Cardiovascular disease (CVD) PRS-330. **C** Obesity PRS-97. **D** Type 2 diabetes (T2D) PRS-46. **E** Depression PRS-8. The number after PRS- indicates the number of SNPs used to calculate the PRS. Dotted vertical lines from left to right represent the quantile of 10%, 25%, 75%, 80%, 85% and 90% in PRS distribution. PRS, polygenic risk score; SNP, single nucleotide polymorphism.



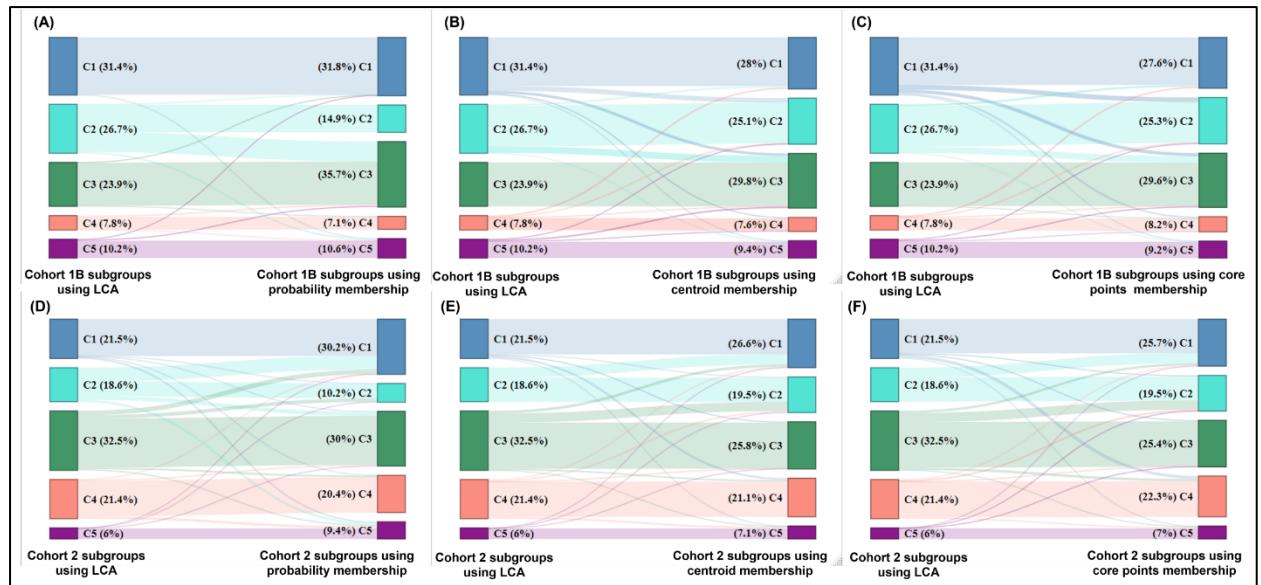
Supplementary Fig. 4. Liver stiffness (LS) levels varied by subtype in the two cohorts. A MCB cohort. LCA-identified subtypes in the development cohort were combined with centroid-assigned subtypes in the validation cohort. **B Tapestry cohort.** Subtypes were assigned using centroid based method. As shown on x-axis, LS was divided into three categories: < 2.5, 2.5–5.0 and > 5.0 kPa. The values in plots represent the number of patients in each subtype with LS levels falling into one of the three LS categories and the percentage over total patients in the subtype (in parentheses). Cohort 1, MCB cohort; Cohort 2, Tapestry cohort; LCA, latent class analysis.



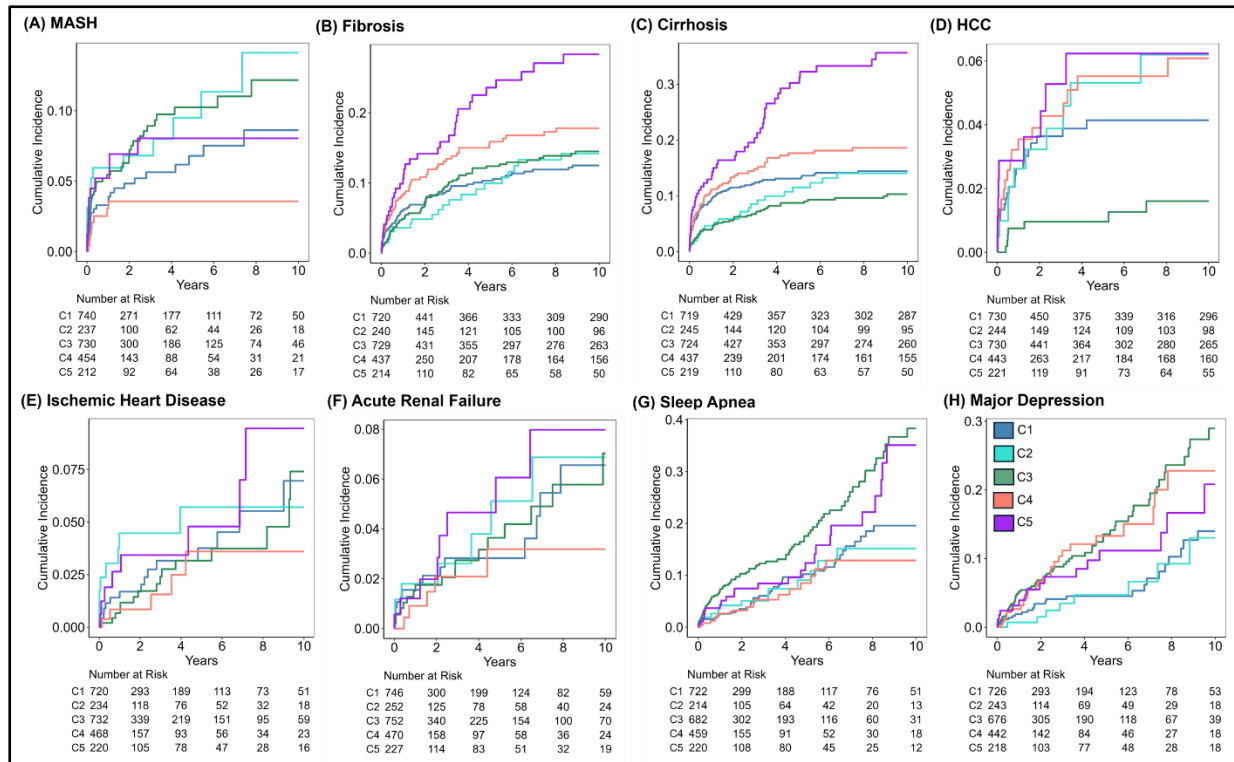
Supplementary Fig. 5. Reproducibility of probability- and core points-based subtype assignments in the MCB validation cohort. A,B Characteristics of five clinical biomarkers in validation cohort and similarity matrix between development and validation cohort using probability-based (MCB 1B, $n = 510$; MCB control, $n = 1,245$) (A) and core points-based subtype assignment methods (MCB 1B, $n = 510$; MCB control, $n = 1,245$) (B). In the boxplots, the lower and upper whiskers represent minimum and maximum, respectively, the center of the box represents median, the box edges represent the first and third quartiles, and the circles denote outliers. Statistical comparisons were performed using two-sided Dunn's test, all with Bonferroni adjustment, and significance was denoted as: **** $P < 0.0001$, *** $P < 0.001$, ** $P < 0.01$, * $P < 0.05$, and ns for not significant. The P values are provided in Supplementary Data 9. The values shown in the heatmap of similarity matrix represent Euclidean distances. Cohort 1A, MCB development cohort; Cohort 1B, MCB validation cohort. C,D Prevalence of 11 comorbidities across control and subtypes assigned using probability- (C) and core points-based methods (D). Error bars indicate 95% confidence intervals. ALT, alanine aminotransferase; BMI, body mass index; HbA1c, hemoglobin A1C; HDL, high-density lipoprotein; MetS, metabolic syndrome; OA, osteoarthritis; T2D, type 2 diabetes.



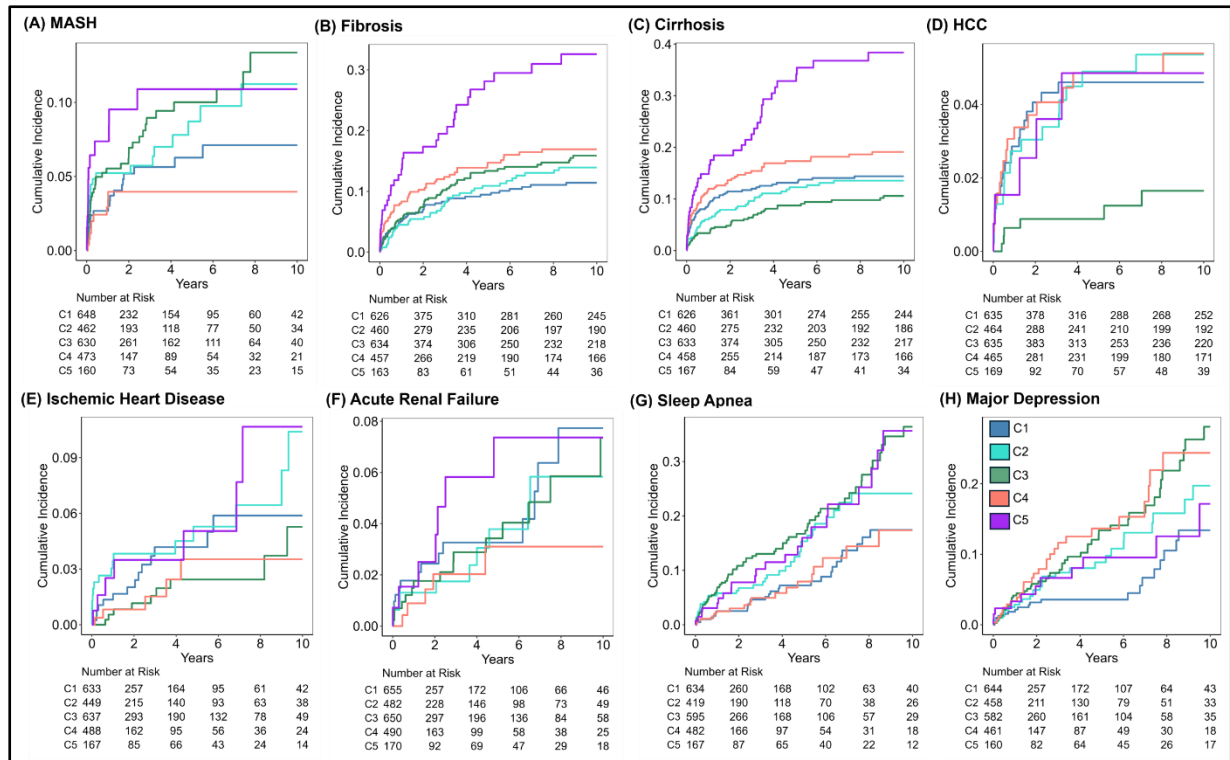
Supplementary Fig. 6. Reproducibility of probability- and core points-based subtype assignments in the Tapestry validation cohort. A,B Characteristics of five clinical biomarkers in validation cohort and similarity matrix between MCB development cohort and Tapestry validation cohort using probability-based (Tapestry, $n = 2,560$; Tapestry control, $n = 1,421$) (A) and core points-based methods (Tapestry, $n = 2,560$; Tapestry control, $n = 1,421$) (B). In the boxplots, the lower and upper whiskers represent minimum and maximum, respectively, the center of the box represents median, the box edges represent the first and third quartiles, and the circles denote outliers. Statistical comparisons were performed using two-sided Dunn's test, all with Bonferroni adjustment, and significance was denoted as: **** $P < 0.0001$, *** $P < 0.001$, ** $P < 0.01$, * $P < 0.05$, and ns for not significant. The P values are provided in Supplementary Data 14. The values shown in the heatmap of similarity matrix represent Euclidean distances. Cohort 1A, MCB development cohort; Cohort 2, Tapestry validation cohort. C,D Prevalence of 11 comorbidities in the Tapestry validation cohort using probability- (C) and core points-based methods (D). Error bars indicate 95% confidence intervals. ALT, alanine aminotransferase; BMI, body mass index; HbA1c, hemoglobin A1c; HDL, high-density lipoprotein; MetS, metabolic syndrome; OA, osteoarthritis; T2D, type 2 diabetes.



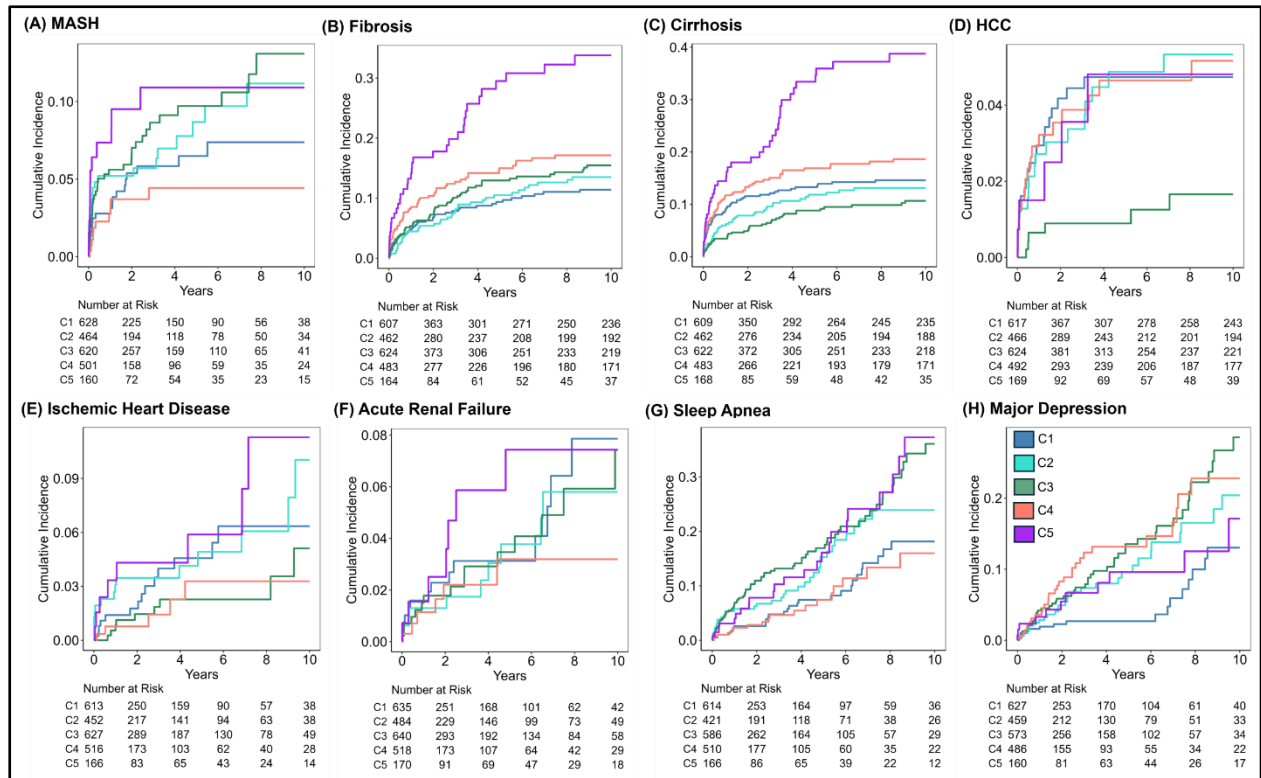
Supplementary Fig. 7. LCA-identified subtypes vs. subtypes assigned in the two validation cohorts. A–C LCA vs. probability (A), centroid (B) and core points (C) based assignments in the MCB validation cohort. D–F LCA vs. probability (D), centroid (E) and core points (F) based assignments in the Tapestry validation cohort. Numbers in parentheses represent the percentage of cases belonging to a subtype based on LCA or assignment approach. Cohort 1B, MCB validation cohort; Cohort 2, Tapestry validation cohort; LCA, latent class analysis.



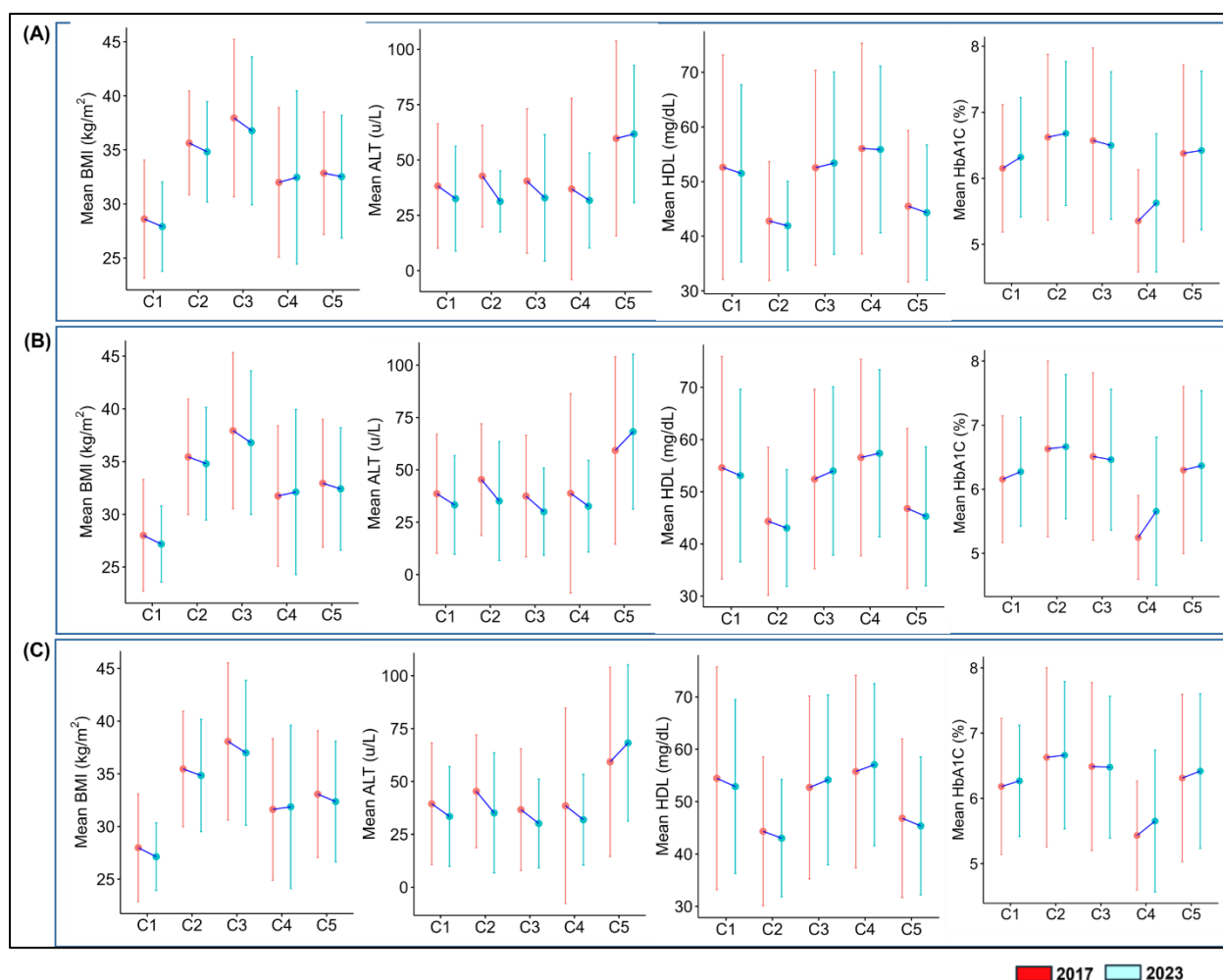
Supplementary Fig. 8. The cumulative risks of developing intrahepatic and extrahepatic complications stratified by subtype based on probability-based assignment in Tapestry cohort. **A–D** Subgroup-specific cumulative incidence for intrahepatic complications. Intrahepatic complications include MASH (**A**), liver fibrosis (**B**), cirrhosis (**C**), and HCC (**D**). **E–H** Subgroup-specific cumulative incidence for extrahepatic complications. Extrahepatic complications include ischemic heart disease (**E**), acute renal failure (**F**), obstructive sleep apnea (**G**), and major depressive disorder (**H**). In subgroups C2–C5, hazard ratio (HR) with 95% confidence interval (CI) were calculated using Cox proportional hazards regression model, with subgroup C1 as the reference (“ref”). Two-sided *P* values were obtained from Wald tests. The initial number at risk included only patients with available follow-up data and no prior relevant diagnoses. HCC, hepatocellular carcinoma; MASH, metabolic dysfunction-associated steatohepatitis.



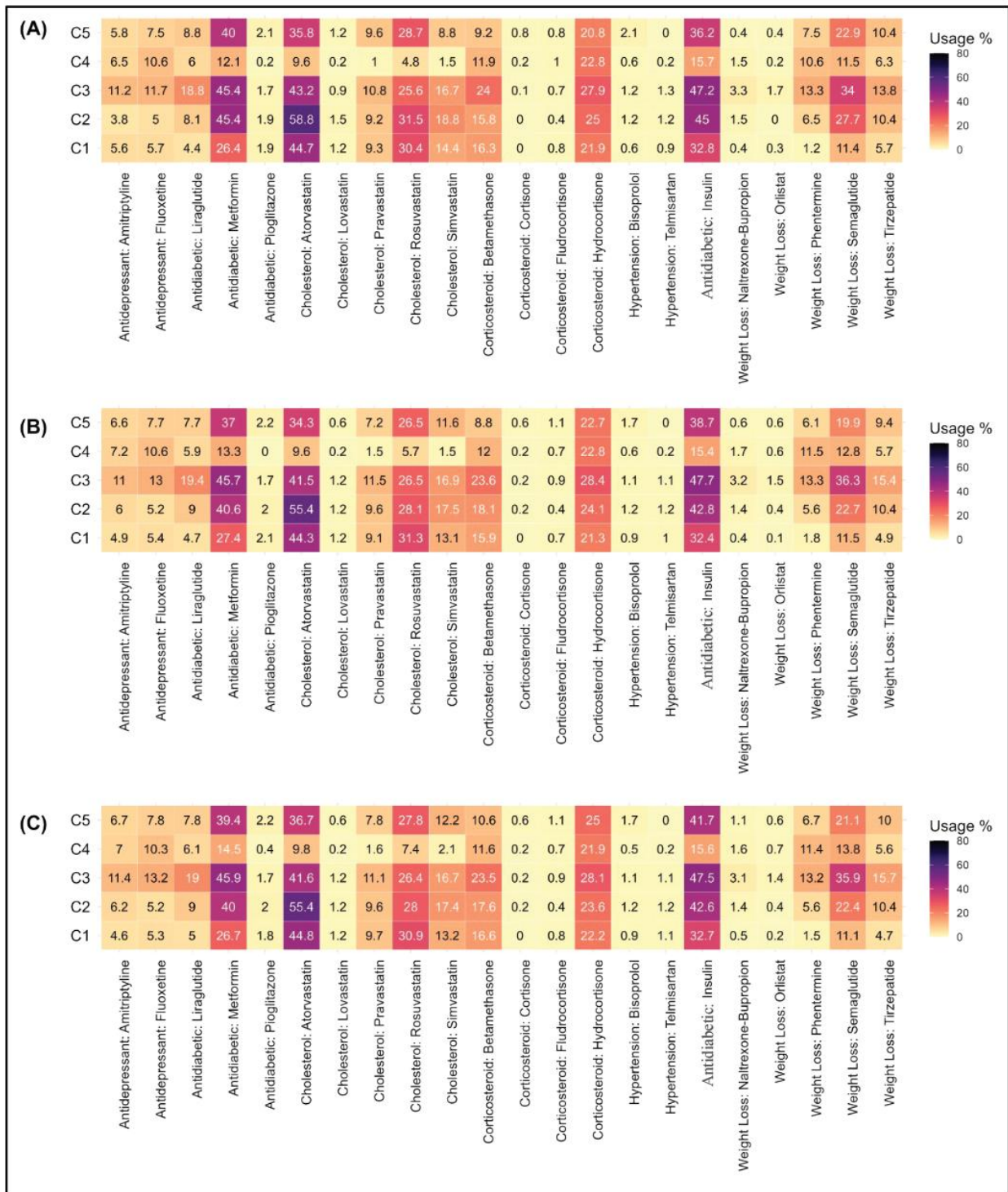
Supplementary Fig. 9. The cumulative risks of developing intrahepatic and extrahepatic complications stratified by subtype from centroid-based assignment in Tapestry cohort. A–D Subgroup-specific cumulative incidence for intrahepatic complications. Intrahepatic complications include MASH (A), liver fibrosis (B), cirrhosis (C), and HCC (D). E–H Subgroup-specific cumulative incidence for extrahepatic complications. Extrahepatic complications include ischemic heart disease (E), acute renal failure (F), obstructive sleep apnea (G), and major depressive disorder (H). In subgroups C2–C5, hazard ratio (HR) with 95% confidence interval (CI) were calculated using Cox proportional hazards regression model, with subgroup C1 as the reference (“ref”). Two-sided *P* values were obtained from Wald tests. The initial number at risk included only patients with available follow-up data and no prior relevant diagnoses. HCC, hepatocellular carcinoma; MASH, metabolic dysfunction-associated steatohepatitis.



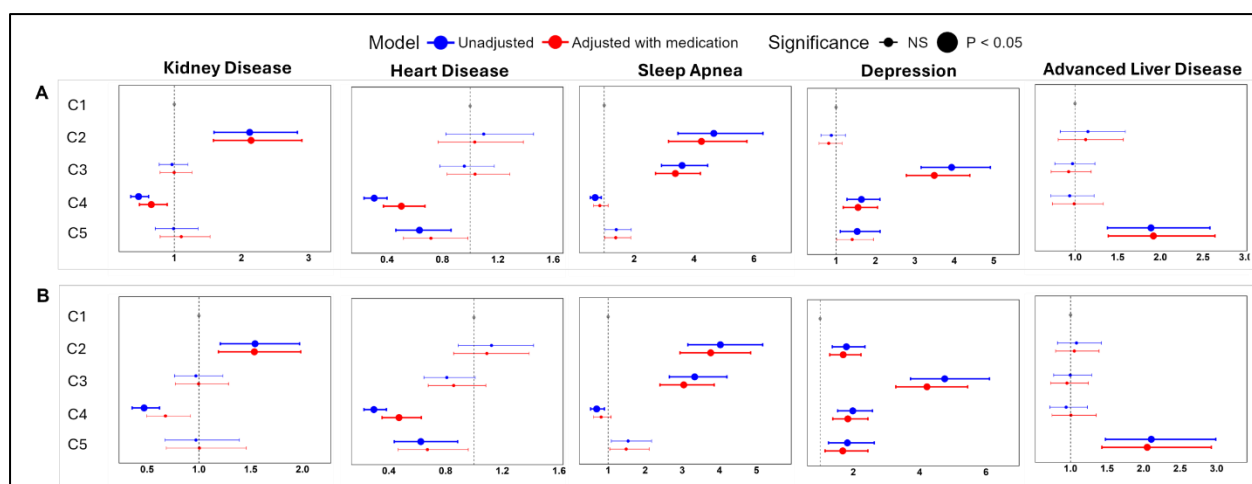
Supplementary Fig. 10. The cumulative risks of developing intrahepatic and extrahepatic complications stratified by subtype from core points-based assignment in Tapestry cohort. A–D Subgroup-specific cumulative incidence for intrahepatic complications. Intrahepatic complications include MASH (A), liver fibrosis (B), cirrhosis (C), and HCC (D). E–H Subgroup-specific cumulative incidence for extrahepatic complications. Extrahepatic complications include ischemic heart disease (E), acute renal failure (F), obstructive sleep apnea (G), and major depressive disorder (H). In subgroups C2–C5, hazard ratio (HR) with 95% confidence interval (CI) were calculated using Cox proportional hazards regression model, with subgroup C1 as the reference (“ref”). Two-sided *P* values were obtained from Wald tests. The initial number at risk included only patients with available follow-up data and no prior relevant diagnoses. HCC, hepatocellular carcinoma; MASH, metabolic dysfunction-associated steatohepatitis.



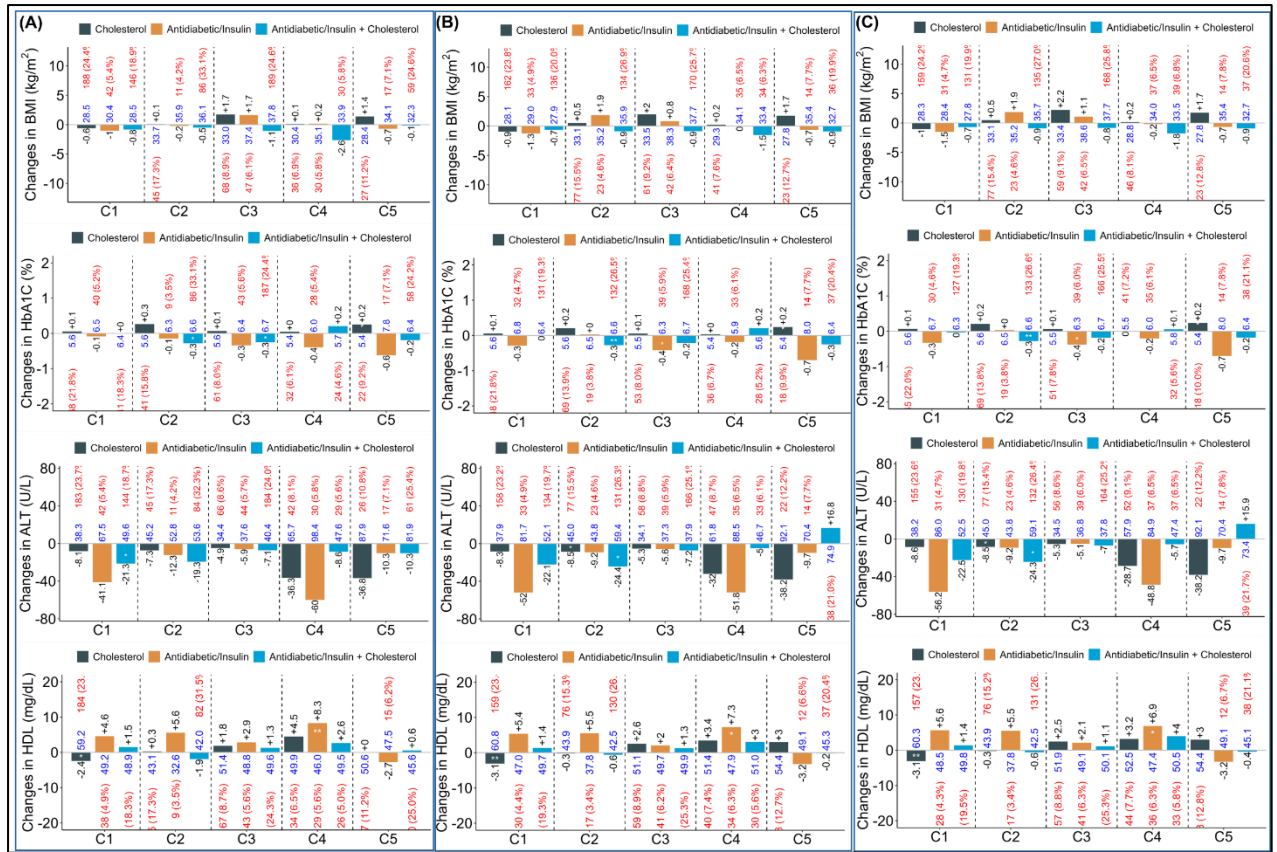
Supplementary Fig. 11. Changes of clinical biomarkers over 6-year follow-up in Tapestry cohort. A–C Subtypes assigned using probability (A), centroid (B), and core points-based methods (C). For each clinical variable (BMI, $n = 1,111$; ALT, $n = 542$; HDL, $n = 615$; HbA1c, $n = 590$), only patients with measurements available in both 2017 and 2023 were included. If a patient has two or more measurements in either year, the average value was used. Error bar represents the standard deviation, and mean values from 2017 and those from 2023 were connected with blue solid lines to indicate overall changes over time. BMI, body mass index; ALT, alanine aminotransferase; HDL, high-density lipoprotein; HbA1c, hemoglobin A1C.



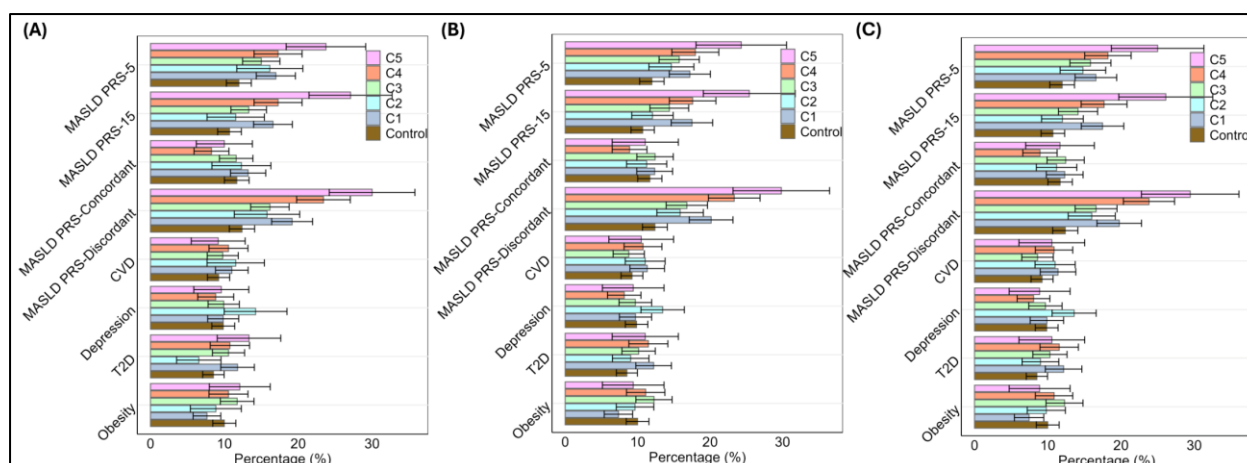
Supplementary Fig. 12. Medication across subtypes in Tapestry cohort. A–C Subtypes assigned using probability (A), centroid (B), and core points-based methods (C). The value represents percent of patients taking the indicated medication represents the percentage. Medications were grouped into classes based on the action of mechanism.



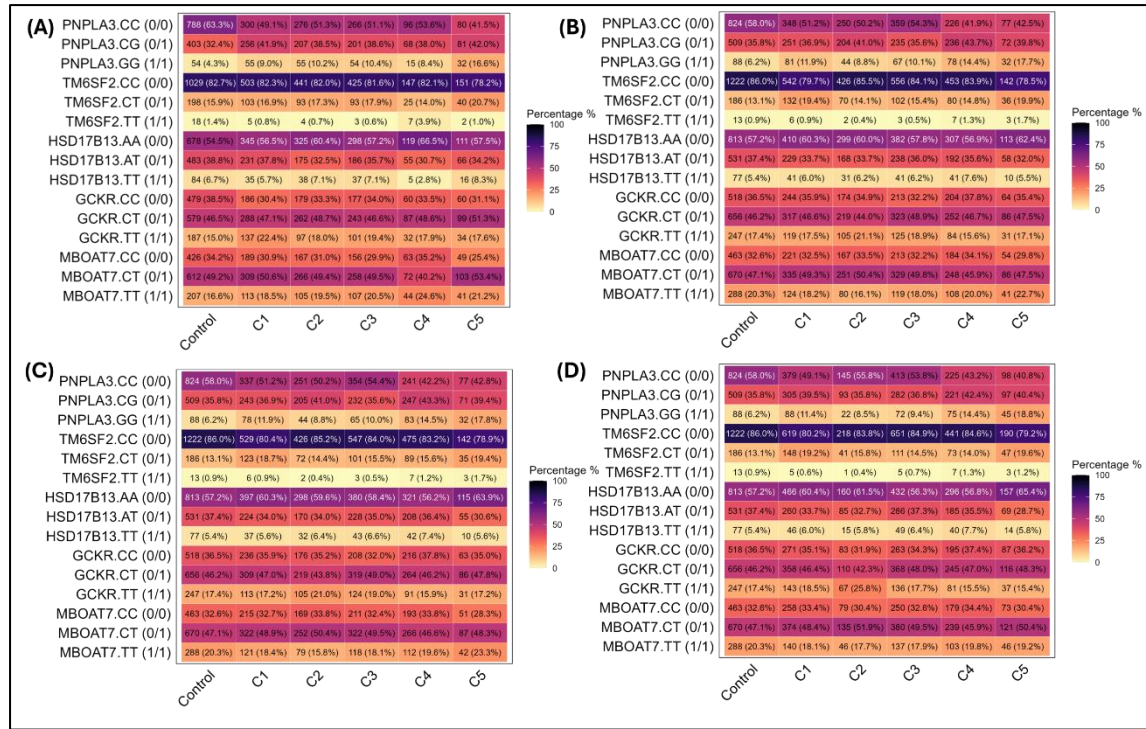
Supplementary Fig. 13. Association of subtypes with medication stratified by liver disease and comorbidities in Tapestry cohort. A,B Subtype assigned with probability (A) and core-points based method (B). For each of subgroups C2–C5, odds Ratio (OR) with 95% Confidence Interval (95% CI) and P value were calculated relative to subgroup C1 using binomial regression with Bonferroni adjustment. X-axis represents OR (95% CI). Exact P values are provided in Supplementary Data 18. Advanced liver disease includes metabolic dysfunction-associated steatohepatitis (MASH), fibrosis, cirrhosis, and hepatocellular carcinoma (HCC). NS, not significant ($P \geq 0.05$).



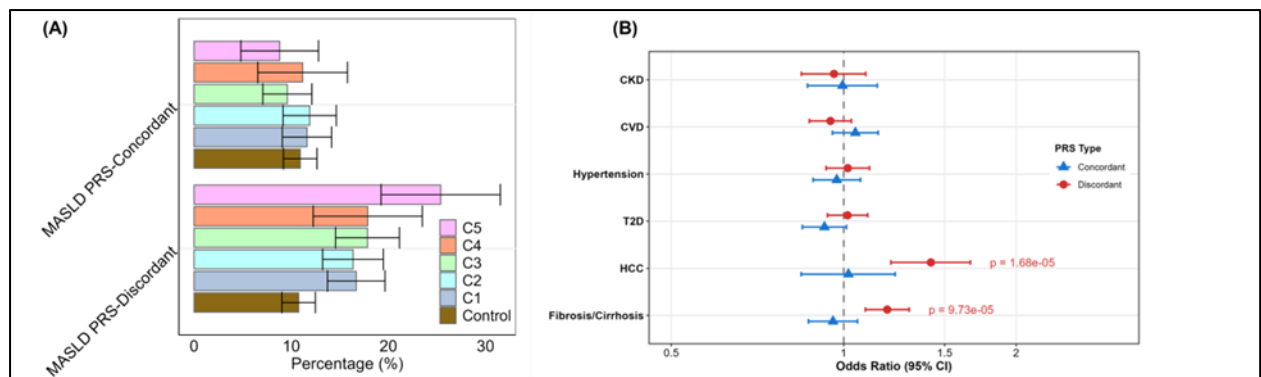
Supplementary Fig. 14. The changes of four clinical biomarkers upon medication in Tapestry cohort. Subtypes were assigned using probability (A), centroid (B), and core points-based method (C). The numbers in red denote the number and proportion (in parentheses) of patients taking the indicated drug. For each clinical variable, values in blue indicate the mean of the first measurement during 2017-2023, and values in black represent the change during 2017-2023 between the mean of the last measurement and the mean of the first measurement. Significant test used two-sided paired t-tests without multiple testing correction, shown as **** $P < 0.0001$, *** $P < 0.001$, ** $P < 0.01$, and * $P < 0.05$ if significant, or non-significance otherwise ($P \geq 0.05$). Exact P values are provided in Supplementary Data 19. ALT, alanine aminotransferase; Antidiabetic/Insulin, insulin belongs to antidiabetic; BMI, body mass index; HbA1c, hemoglobin A1C; HDL, high-density lipoprotein.



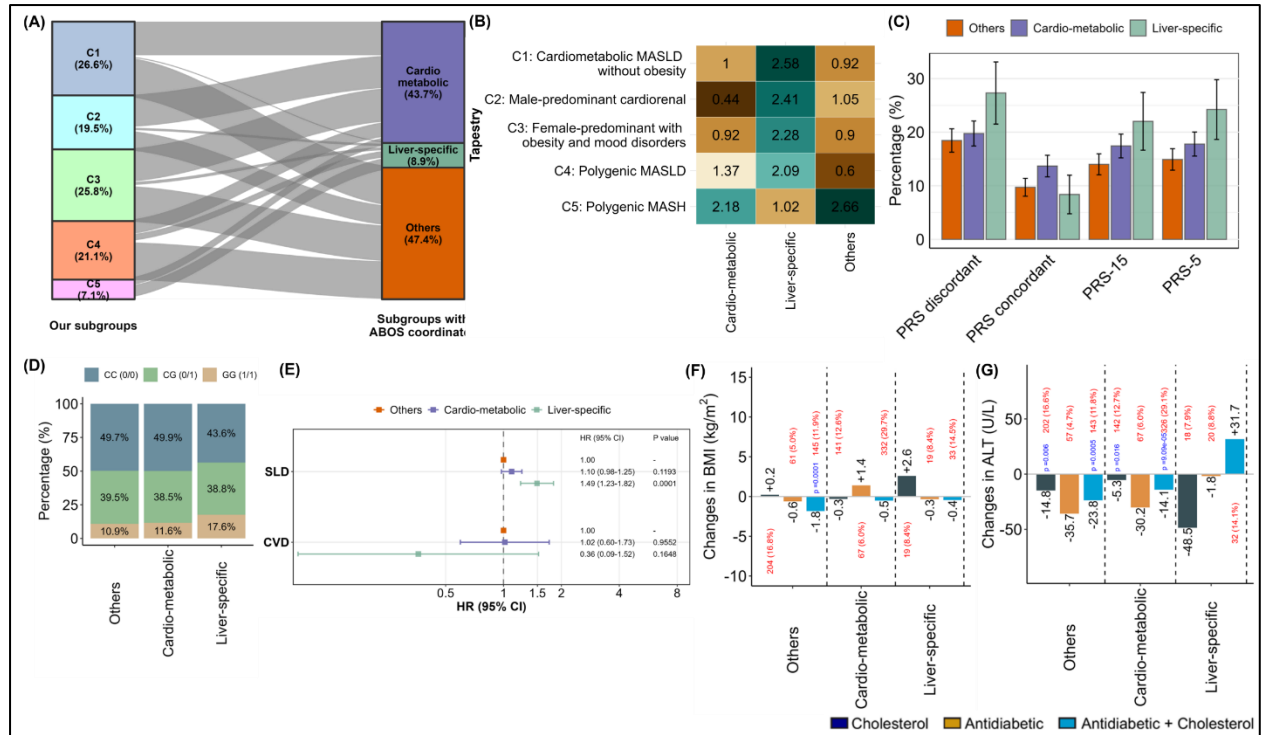
Supplementary Fig. 15. Proportion of subjects with top decile of PRSs in the assigned subtypes and control in Tapestry cohort. A Probability-based method. **B** Centroid-based method. **C** Core points-based method. The eight PRSs were MASLD PRS-5, PRS-15, concordant PRS (2 SNPs in *GCKR* and *MTTP*), and discordant PRS (3 SNPs in *PNPLA3*, *TM6SF2* and *APOE*), as well as type 2 diabetes (T2D) PRS-46, obesity PRS-97, depression PRS-8, and cardiovascular disease (CVD) PRS-330. The number after PRS- indicates the number of SNPs used in the PRS calculation. Error bars indicate 95% confidence intervals.



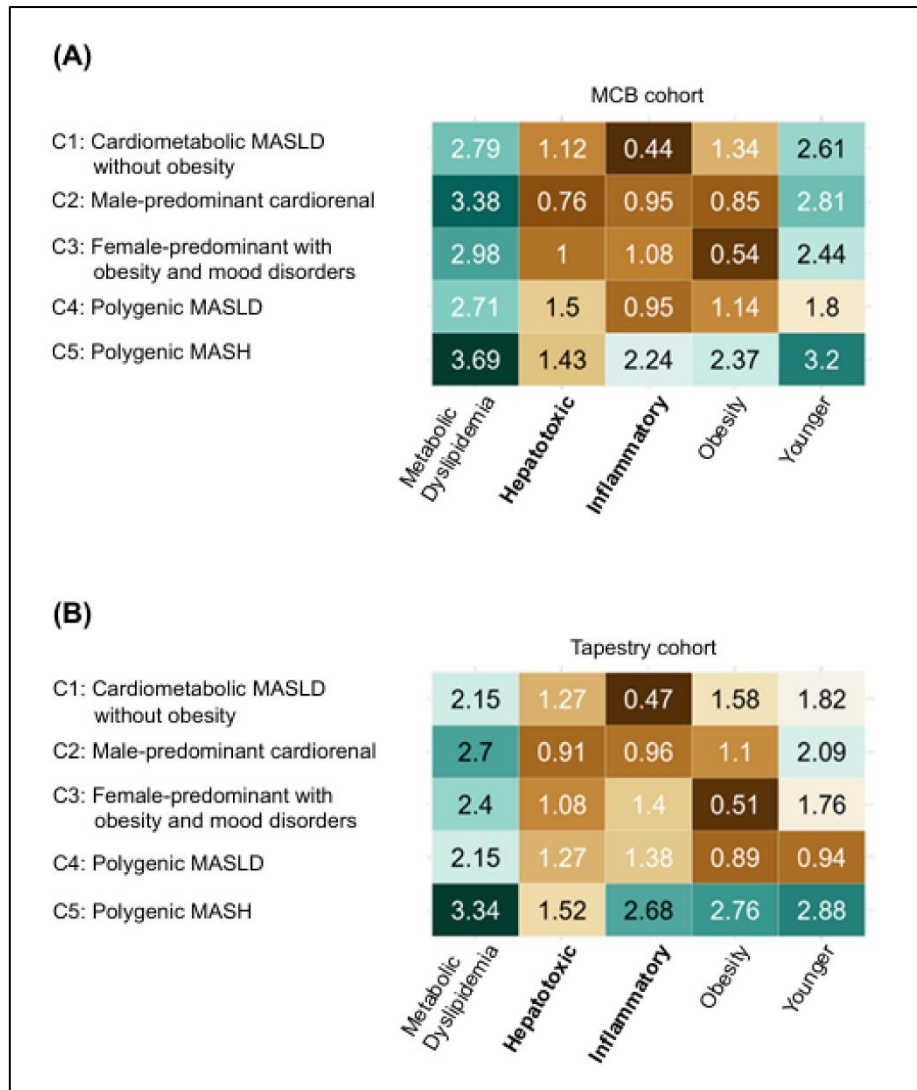
Supplementary Fig. 16. Enrichment of key MASLD genetic variants across subtypes and control. **A** Subtypes in the MCB cohort. LCA-identified subtypes in the development cohort were combined with centroid-assigned subtypes in the validation cohort. **B–D** Subtypes in Tapestry cohort assigned via centroid (**B**), probability (**C**), and core points-based methods (**D**). Risk variants rs738409 C>G in *PNPLA3*, rs58542926 C>T in *TM6SF2*, rs1260326 C>T in *GCKR*, and rs641738 C>T in *MBOAT7*, as well as protective variant rs13118664 A>T in *HSD17B13* were assessed. For control and each subtype, the number of carriers per genotype was shown, along with the percentage of carriers over total individuals in parentheses.



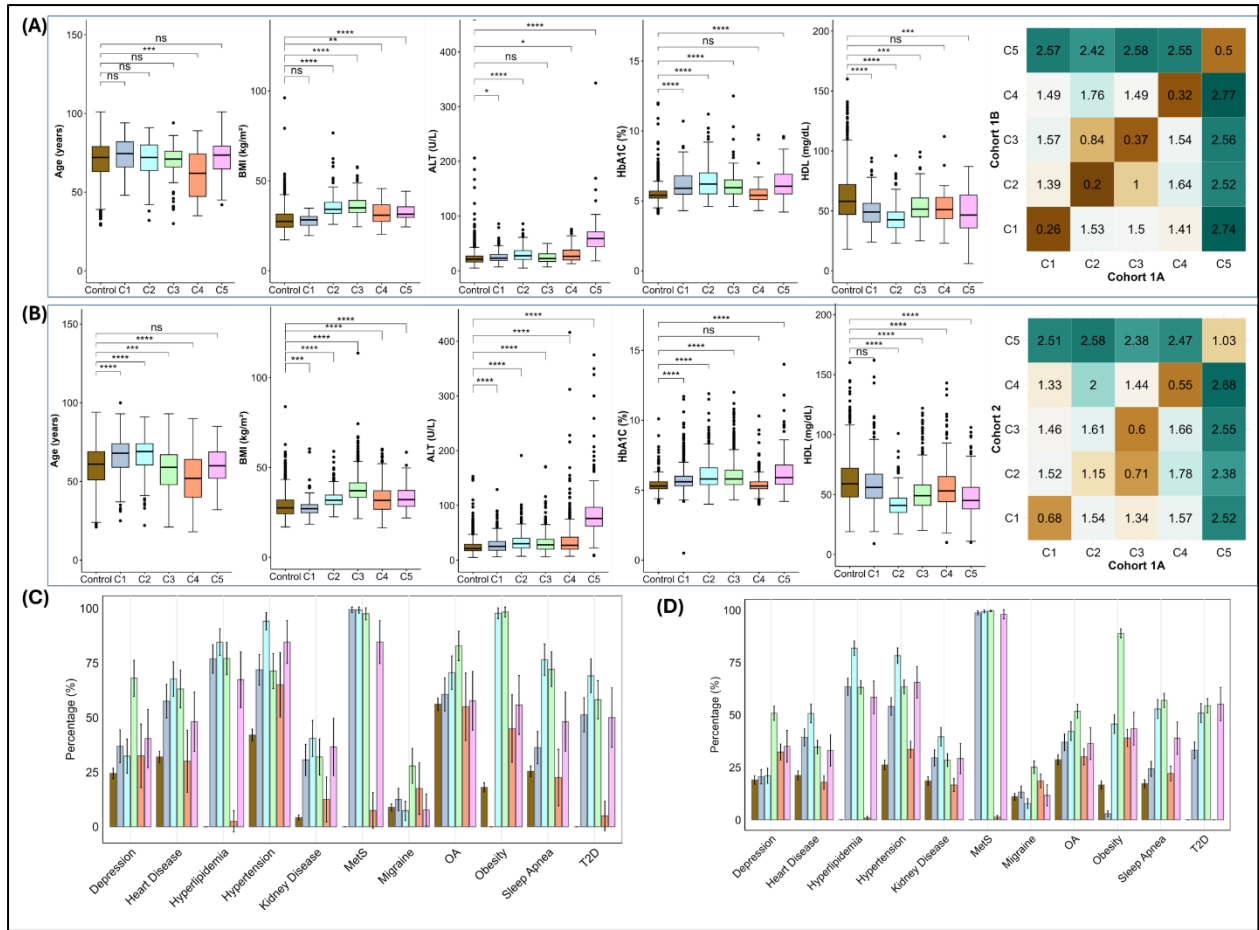
Supplementary Fig. 17. MASLD partitioned PRSs and the associations with liver and cardiometabolic diseases in the MCB cohort. (A) Percentages of patients in the top decile of partitioned PRSs. The LCA-identified subtypes in the development cohort were combined with centroid-assigned subtypes in the validation cohort. Concordant PRS uses 2 SNPs in *GCKR* and *MTTP* and discordant PRS uses 3 SNPs in *PNPLA3*, *TM6SF2* and *APOE*. Error bars indicate 95% confidence interval (CI). (B) Association between discordant and concordant PRS with liver and cardiometabolic diseases. Binomial regression was used to estimate odds ratio with 95% CI and *P* value. CVD, cardiovascular disease; CKD, chronic kidney disease; HCC, hepatocellular carcinoma; T2D, type 2 diabetes.



Supplementary Fig. 18. Cross-study comparison of two subtyping approaches in Tapestry. (A) Correspondence between subgroups assigned with centroid based method and those assigned with ABOS coordinates in the Tapestry cohort. Raverdy et al. (2024) identified six clusters (1–6) in the ABOS cohort (pmid: 39653777). Two of them were named as Cardio-metabolic cluster (Cluster 2) and Liver-specific cluster (Cluster 5) in Raverdy et al. study, while the other four clusters (1, 3, 4 and 6 were combined into a “control” cluster by Raverdy et al.), collectively labeled as "Others" here, had lower severity of MASLD. Subgroups in Tapestry were assigned as "Cardio-metabolic", "Liver-specific", and "Others" based on the coordinates in the ABOS cohort. (B) Heatmap of similarity matrix between centroid and ABOS coordinate based subgroups in Tapestry cohort. The value in the heatmap represents Euclidean distance. (C) Proportion of subjects with top decile of MASLD PRSs in the subgroups in Tapestry cohort ($n = 2,560$). The four PRSs were MASLD PRS-5, PRS-15, concordant PRS (2 SNPs in *GCKR* and *MTTP*), and discordant PRS (3 SNPs in *PNPLA3*, *TM6SF2* and *APOE*). Error bars indicate 95% confidence interval (CI). (D) Proportion of subjects in each subgroup carrying different genotypes at rs738409 (*PNPLA3*) in Tapestry cohort. (E) Cumulative incidence of severe liver disease ($n = 1,491$) and cardiovascular disease (CVD) ($n = 2,360$) in Tapestry cohort. For Cardio-metabolic and Liver-specific clusters, Cox regression was used to estimate hazard ratio (HR) and 95% CI relative to the “Others” cluster. Two-sided P values were obtained from Wald tests. Severe liver disease (SLD) includes metabolic dysfunction-associated steatohepatitis (MASH), fibrosis, cirrhosis, hepatocellular carcinoma (HCC), where cases of liver transplantation were also included. (F–G) Changes of BMI (F) and ALT (G) upon medication in Tapestry cohort. Red numbers indicate the number and proportion (in parentheses) of patients taking the indicated drug. Black values represent the change. Statistical significance was assessed using two-sided paired t-tests, with exact P values reported where significant. In C–G, subgroups were assigned on the basis of ABOS coordinates. ALT, alanine aminotransferase; BMI, body mass index.



Supplementary Fig. 19. Heatmaps of similarity matrix between two sets of subgroups in MCB cohort (A) and Tapestry cohort (B). For C1–C5 in MCB (A), LCA-identified subgroups in the development cohort were combined with centroid-assigned subtypes in the validation cohort. For C1–C5 in Tapestry (B), subgroups were assigned with centroid based method. Hong et al., (2025) identified five MASLD subgroups in UK Biobank training cohort, with the Hepatotoxic (high risk of severe liver disease) and Inflammatory (highest extrahepatic complications) subgroups together representing the high-risk group (in bold text) and the other three subgroups (Metabolic-Dyslipidemia, Obesity and Younger) representing the low-risk group (PMID: 40640848). These five subgroups were assigned to both MCB and Tapestry using coordinates from UK Biobank training cohort. The value in the heatmap represents Euclidean distance, with shorter distance indicating higher similarity.



Supplementary Fig. 20. Consistency of LCA-derived subgroups across development and validation cohorts. A Characteristics of five clinical biomarkers in MCB validation cohort (MCB 1B, $n = 510$; MCB control, $n = 1,245$) and similarity matrix between MCB development cohort (cohort 1A) and validation cohort (cohort 1B). **B** Characteristics of clinical biomarkers in Tapestry validation cohort (Tapestry, $n = 2,560$; Tapestry control, $n = 1,421$) and similarity matrix between MCB development cohort (cohort 1A) and validation Tapestry cohort. In the boxplots, the lower and upper whiskers represent minimum and maximum, respectively, the center of the box represents median, the box edges represent the first and third quartiles, and the circles denote outliers. Some outliers were not shown in the plots. Statistical comparisons were performed using two-sided Dunn's test, all with Bonferroni adjustment, and significance was denoted as: **** $P < 0.0001$, *** $P < 0.001$, ** $P < 0.01$, * $P < 0.05$, and ns for not significant. The exact P values are provided in Source Data. The values shown in the heatmap of similarity matrix represent Euclidean distances (the lower the better). Cohort 1A, MCB development cohort; Cohort 1B, MCB validation cohort; Cohort 2, Tapestry validation cohort. **C,D** Prevalence of 11 comorbidities in the MCB validation cohort (**C**) and Tapestry validation cohort (**D**). Error bars indicate 95% confidence intervals. ALT, alanine aminotransferase; BMI, body mass index; HbA1c, hemoglobin A1c; HDL, high-density lipoprotein; MetS, metabolic syndrome; OA, osteoarthritis; T2D, type 2 diabetes.

Supplementary Table 1: Presence of cardiometabolic risk factors in MASLD patients

Cardiometabolic risk factors (CMRFs)	MCB cohort		Tapestry Study	
	Case	Control	Case	Control
	(N=2042)	(N=1245)	(N=2560)	(N=1421)
BMI ≥ 25 kg/m ²	1793 (87.8%)	833 (66.9%)	2288 (89.4%)	963 (67.8%)
HbA1c $\geq 5.7\%$ or type 2 diabetes or treatment for type 2 diabetes	1612 (78.9%)	441 (35.4%)	1810 (70.7%)	441 (31%)
Presence of hypertension or antihypertensive drug treatment	1599 (78.3%)	523 (42%)	1483 (57.9%)	372 (26.2%)
Plasma triglycerides ≥ 150 mg/dL or lipid lowering treatment	798 (39.1%)	200 (16.1%)	970 (37.9%)	222 (15.6%)
Plasma HDL-cholesterol (≤ 40 mg/dL (Male); ≤ 50 mg/dL (Female)) or lipid lowering treatment)	930 (45.5%)	234 (18.8%)	1067 (41.7%)	260 (18.3%)
Number of CMRFs present				
0	19 (0.9%)	190 (15.3%)	57 (2.2%)	269 (18.9%)
≥ 1	2023 (99.1%)	1055 (84.7%)	2503 (97.8%)	1152 (81.1%)
≥ 2	1904 (93.2%)	706 (56.7%)	2260 (88.3%)	712 (50.1%)
≥ 3	1558 (76.3%)	331 (26.6%)	1678 (65.5%)	296 (20.8%)
≥ 4	922 (45.2%)	117 (9.4%)	875 (34.2%)	82 (5.8%)
5	325 (15.9%)	22 (1.8%)	302 (11.8%)	16 (1.1%)

MASLD cases were defined using the criteria prior to the release of the new fatty liver disease nomenclature in 2023.

BMI, body mass index; HbA1c, hemoglobin A1C; HDL, high-density lipoprotein

Supplementary Table 2: Criteria used for continuous to categorical variable conversion

Variable	Range	Category	Variable	Range	Category
BMI (kg/m ²)	<18.5	Underweight	AST (U/L)	>48	High
	18.5–24.9	Normal		8 to 48	Normal
	25–29.9	Overweight		<8	Low
	30–34.9	Obesity I	HbA1c (%)	≥6.5	Diabetes
	35–39.9/ ≥40	Obesity II/III		5.7 to 6.4	Prediabetes
LDL (mg/dL)	<70	Best		<5.7	Normal
	70 to 128	Optimal	ALT (U/L)	>55	High
	≥129	High		7 to 55	Normal
CRP (mg/L)	<8	Low	BUN (mg/dL)	<6	Low
	≥8	High	PT (seconds)	>12.5	High
TG (mg/dL)	≥150	High		9.4 to 12.5	Normal
	<150	Desirable		<9.4	Low
ALP (U/L)	>129	High	PRS	≥ quantile (0.9)	High
	40 to 129	Normal		≥quantile (0.4) or <quantile (0.9)	Intermediate
	<40	Low		<quantile (0.4)	Low

BMI, body mass index; LDL, low-density lipoprotein; HDL, high-density lipoprotein; CRP, c-reactive protein; TG, triglycerides; AST, aspartate aminotransferase; HbA1c, hemoglobin A1c; ALT, alanine aminotransferase; BUN, blood urea nitrogen; PT, prothrombin time; PRS, polygenic risk score; ALP, alkaline phosphatase.

Supplementary Table 3: LCA model-fit indices in the MCB development cohort

No of class	Log-likelihood	AIC	BIC	SABIC	Entropy	Average(Posterior)	Min(proportion)	PLMR
1	-12509.07	25056.1	25157.5	25097.1	-	1	100	-
2	-12074.25	24226.5	24434.5	24310.7	0.926	0.9847	43.6684	2.47E-163
3	-11911.35	23940.7	24255.4	24068	0.892	0.9615	10.4439	3.37E-54
4	-11766.75	23691.5	24112.9	23861.9	0.801	0.8913	9.6606	4.66E-47
5	-11666.56	23531.1	24059.2	23744.7	0.81	0.8896	9.1384	4.95E-30
6	-11617.64	23473.3	24108.1	23730	0.796	0.8704	4.6345	1.75E-11
7	-11577.96	23433.9	24175.4	23733.8	0.786	0.8411	4.3734	1.91E-08
8	-11541.33	23400.7	24248.8	23743.7	0.805	0.8500	3.4595	1.77E-07
9	-11511.16	23380.3	24335.2	23766.5	0.791	0.8400	3.3943	1.60E-05
10	-11482.95	23363.9	24425.4	23793.3	0.722	0.7850	3.1984	5.87E-05

AIC, Akaike Information Criterion; BIC, Bayesian Information Criterion, SABIC, sample-size adjusted Bayesian Information Criterion; Average_(Posterior), average posterior probability; Min_(proportion), minimum class Proportion; PLMR, value of one-sided Lo-Mendell–Rubin likelihood ratio test.

Supplementary Table 4: Baseline characteristics of subgroups in MCB development cohort based on data within one year before and after first diagnosis of MASLD

Variable	Metric	C1	C2	C3	C4	C5
		N [mean(SD)]	N[mean(SD)]	N[mean(SD)]	N[mean(SD)]	N[mean(SD)]
HbA1c	After diagnosis	284 [6.10 (1.34)]	277 [6.39 (1.51)]	247 [6.51 (1.78)]	56 [5.47 (0.62)]	80 [6.03 (1.20)]
	Before diagnosis	213 [6.22 (1.34)]	239 [6.50 (1.55)]	227 [6.45 (1.40)]	31 [5.39 (0.64)]	62 [6.41 (1.69)]
	Average	338 [6.39 (1.16)]	332 [6.62 (1.24)]	300 [6.52 (1.36)]	74 [5.50 (0.80)]	100 [6.38 (1.28)]
ALP	After diagnosis	157 [106.75 (81.68)]	142 [99.80 (116.77)]	151 [105.01 (77.63)]	50 [85.94 (33.84)]	59 [121.42 (175.09)]
	Before diagnosis	120 [111.05 (131.60)]	101 [95.33 (97.94)]	125 [113.17 (103.34)]	37 [100.03 (48.84)]	41 [126.56 (135.32)]
	Average	178 [92.41 (40.60)]	166 [93.40 (53.58)]	178 [91.84 (43.17)]	59 [89.05 (34.33)]	63 [218.60 (270.62)]
ALT	After diagnosis	123 [36.88 (24.33)]	117 [43.85 (41.93)]	107 [45.57 (90.55)]	40 [34.30 (17.63)]	49 [69.76 (95.82)]
	Before diagnosis	90 [44.58 (55.23)]	77 [71.60 (255.71)]	88 [39.66 (39.68)]	27 [41.11 (24.55)]	29 [73.34 (67.64)]
	Average	138 [27.10 (13.57)]	138 [30.38 (15.96)]	128 [28.16 (19.18)]	46 [30.76 (14.72)]	52 [99.63 (187.12)]
AST	After diagnosis	142 [37.30 (32.97)]	127 [41.98 (34.25)]	123 [41.47 (62.40)]	46 [31.93 (27.44)]	47 [54.19 (33.95)]
	Before diagnosis	101 [47.83 (71.53)]	85 [55.60 (161.21)]	107 [33.43 (22.69)]	30 [39.70 (34.94)]	31 [59.58 (42.13)]
	Average	159 [27.58 (10.07)]	151 [29.76 (19.61)]	149 [26.05 (9.05)]	52 [28.44 (12.36)]	51 [137.65 (163.08)]
BMI	After diagnosis	229 [31.89 (23.39)]	169 [34.83 (5.71)]	157 [39.83 (23.68)]	55 [33.45 (9.72)]	64 [32.30 (6.12)]
	Before diagnosis	166 [28.80 (6.46)]	114 [35.18 (6.07)]	126 [37.11 (6.49)]	48 [32.90 (9.04)]	36 [32.62 (5.46)]
	Average	247 [28.04 (4.29)]	178 [35.36 (5.59)]	169 [36.96 (6.13)]	61 [33.33 (9.20)]	66 [34.38 (21.85)]
HDL	After diagnosis	97 [101.18 (55.31)]	82 [74.76 (49.42)]	80 [93.35 (52.92)]	23 [74.13 (30.49)]	26 [110.54 (81.04)]
	Before diagnosis	91 [105.19 (55.96)]	77 [79.16 (51.43)]	82 [97.16 (54.57)]	16 [77.31 (34.29)]	20 [113.35 (56.88)]
	Average	130 [50.78 (15.58)]	112 [41.07 (11.14)]	114 [51.76 (11.93)]	30 [51.70 (14.92)]	36 [46.72 (16.97)]
LDL	After diagnosis	286 [119.22 (39.31)]	284 [119.38 (39.53)]	280 [120.29 (37.19)]	72 [100.08 (27.87)]	85 [120.73 (45.15)]
	Before diagnosis	272 [118.83 (39.91)]	274 [116.97 (38.58)]	295 [121.65 (38.98)]	58 [102.76 (33.34)]	79 [120.53 (40.16)]
	Average	380 [84.19 (36.10)]	388 [72.16 (31.04)]	372 [87.29 (34.30)]	95 [91.32 (30.06)]	122 [82.28 (38.95)]
TG	After diagnosis	221 [150.34 (113.70)]	145 [174.62 (105.99)]	145 [169.54 (84.10)]	49 [113.35 (56.63)]	54 [140.87 (83.63)]
	Before diagnosis	193 [162.25 (112.27)]	123 [194.02 (109.62)]	135 [175.21 (80.77)]	37 [108.57 (53.94)]	43 [154.91 (82.03)]
	Average	281 [141.53 (88.81)]	183 [164.37 (138.55)]	184 [166.11 (98.28)]	59 [107.53 (49.25)]	71 [147.04 (75.68)]

Data was from within one year before ("-1 to 0") and after MASLD onset ("0 to 1").

ALP, alkaline phosphatase; ALT, alanine aminotransferase; AST, aspartate aminotransferase; BMI, body mass index; HbA1c, hemoglobin A1C; HDL, high-density lipoprotein; LDL, low-density lipoprotein; TG, triglycerides.

Supplementary Table 5: Performance metrics of three subgroup assignment methods in the two validation cohorts

	MCB validation cohort					
	Probability Based		Centroid Based		Core Points Based	
Cluster	Sensitivity	Specificity	Sensitivity	Specificity	Sensitivity	Specificity
C1	0.982	0.997	0.944	0.932	0.936	0.924
C2	1	0.862	0.883	0.94	0.876	0.939
C3	0.659	0.994	0.803	1	0.801	0.997
C4	0.889	0.991	0.872	0.987	0.809	0.987
C5	0.972	0.989	0.938	0.985	0.957	0.985
	Tapestry cohort					
	Probability Based		Probability Based		Probability Based	
Cluster	Sensitivity	Specificity	Sensitivity	Specificity	Sensitivity	Specificity
C1	0.667	0.98	0.724	0.969	0.7249	0.9611
C2	0.788	0.883	0.675	0.933	0.674	0.933
C3	0.892	0.918	0.977	0.902	0.977	0.8973
C4	0.604	0.996	0.913	0.973	0.8739	0.9475
C5	0.95	0.974	0.801	0.996	0.8167	0.9971

The Latent Class Analysis (LCA) identified subgroups were used as the reference in the comparison

Supplementary Table 6: LCA-based vs. assignment-based subgroups in the two validation cohorts

LCA subgroups → membership based assigned subgroups	MCB validation cohort			LCA subgroups → membership based assigned subgroups	Tapestry cohort		
	centroid	probability	core points		centroid	probability	core points
	No. cases (%)	No. cases (%)	No. Cases (%)		No. cases (%)	No. cases (%)	No. cases (%)
C1 → C1	135 (84.38)	132 (82.50)	159 (99.38)	C1 → C1	492 (89.29)	477 (86.57)	515 (93.47)
C1 → C2	13 (8.13)	13 (8.13)	0 (0.00)	C1 → C2	12 (2.18)	13 (2.36)	3 (0.54)
C1 → C3	7 (4.38)	9 (5.63)	0 (0.00)	C1 → C3	11 (2.00)	12 (2.18)	11 (2.00)
C1 → C4	3 (1.88)	5 (3.13)	0 (0.00)	C1 → C4	27 (4.90)	40 (7.26)	3 (0.54)
C1 → C5	2 (1.25)	1 (0.63)	1 (0.63)	C1 → C5	9 (1.63)	9 (1.63)	19 (3.45)
C2 → C1	4 (2.94)	5 (3.68)	1 (0.74)	C2 → C1	139 (29.26)	138 (29.05)	191 (40.21)
C2 → C2	113 (83.09)	113 (83.09)	76 (55.88)	C2 → C2	336 (70.74)	337 (70.95)	205 (43.16)
C2 → C3	0 (0.00)	0 (0.00)	55 (40.44)	C2 → C3	0 (0.00)	0 (0.00)	51 (10.74)
C2 → C4	18 (13.24)	17 (12.50)	0 (0.00)	C2 → C4	0 (0.00)	0 (0.00)	0 (0.00)
C2 → C5	1 (0.74)	1 (0.74)	4 (2.94)	C2 → C5	0 (0.00)	0 (0.00)	28 (5.89)
C3 → C1	0 (0.00)	0 (0.00)	1 (0.82)	C3 → C1	36 (4.33)	30 (3.61)	62 (7.45)
C3 → C2	0 (0.00)	0 (0.00)	0 (0.00)	C3 → C2	128 (15.38)	129 (15.50)	48 (5.77)
C3 → C3	122 (100.00)	121 (99.18)	120 (98.36)	C3 → C3	646 (77.64)	636 (76.44)	684 (82.21)
C3 → C4	0 (0.00)	1 (0.82)	1 (0.82)	C3 → C4	20 (2.40)	32 (3.85)	23 (2.76)
C3 → C5	0 (0.00)	0 (0.00)	0 (0.00)	C3 → C5	2 (0.24)	5 (0.60)	15 (1.80)
C4 → C1	4 (10.00)	4 (10.00)	0 (0.00)	C4 → C1	10 (1.82)	9 (1.64)	1 (0.18)
C4 → C2	1 (2.50)	1 (2.50)	0 (0.00)	C4 → C2	19 (3.47)	20 (3.65)	0 (0.00)
C4 → C3	1 (2.50)	1 (2.50)	4 (10.00)	C4 → C3	1 (0.18)	1 (0.18)	19 (3.47)
C4 → C4	34 (85.00)	34 (85.00)	35 (87.50)	C4 → C4	493 (89.96)	499 (91.06)	495 (90.33)
C4 → C5	0 (0.00)	0 (0.00)	1 (2.50)	C4 → C5	25 (4.56)	19 (3.47)	33 (6.02)
C5 → C1	0 (0.00)	0 (0.00)	1 (1.92)	C5 → C1	3 (1.95)	4 (2.60)	3 (1.95)
C5 → C2	1 (1.92)	2 (3.85)	0 (0.00)	C5 → C2	3 (1.95)	1 (0.65)	4 (2.60)
C5 → C3	4 (7.69)	3 (5.77)	3 (5.77)	C5 → C3	3 (1.95)	2 (1.30)	2 (1.30)
C5 → C4	2 (3.85)	2 (3.85)	0 (0.00)	C5 → C4	0 (0.00)	0 (0.00)	0 (0.00)
C5 → C5	45 (86.54)	45 (86.54)	48 (92.31)	C5 → C5	145 (94.16)	147 (95.45)	145 (94.16)

Number and percentages of cases with the same subgroup by LCA and by assignment are bold

Supplementary Table 7: Accuracy of subgroup assignments in subsets of MASLD patients diagnosed with type 2 diabetes or obesity

Membership assignment	MCB validation cohort			Tapestry Study		
	Whole dataset (N=510)	T2D subset (N=289)	Obese subset (N=365)	Whole dataset (N=2560)	T2D subset (N=1005)	Obese subset (N=1747)
Probability Based	85.8%	84.8%	80.8%	79.8%	77.4%	79.2%
Centroid Based	88.0%	87.5%	86.8%	82.5%	80.6%	83.3%
Core points Based	87.0%	86.9%	86.0%	81.8%	80.1%	82.8%

If a patient had both type 2 diabetes (T2D) and obesity, it was included in both T2D subset and obese subset.

Supplementary Table 8: Risk of MASLD related intrahepatic and extrahepatic disease complications across subtypes in MCB cohort

Outcome	Subtype	Events	Censored	% Events	Overall events (%)	Cox regression model		Competing risk regression model	
						HR (95% CI)	P value	HR (95% CI)	P Value
MASH	C1	31	538	5.45	8.11	1	-	1	-
	C2	42	456	8.43		1.53(0.96-2.43)	0.0726	1.51(0.95-2.39)	0.082
	C3	50	436	10.29		1.85(1.18-2.89)	0.0072	1.86(1.19-2.92)	0.0068
	C4	8	158	4.82		0.86(0.39-1.87)	0.7015	0.95(0.44-2.08)	0.91
	C5	23	157	12.78		2.22(1.30-3.81)	0.0037	2.14(1.25-3.64)	0.0053
Fibrosis	C1	63	498	11.23	13.65	1	-	1	-
	C2	66	417	13.66		1.23(0.87-1.73)	0.2456	0.80(0.42-1.55)	0.52
	C3	59	405	12.72		1.26(0.89-1.80)	0.1982	1.46(0.81-2.61)	0.21
	C4	23	133	14.74		1.47(0.91-2.37)	0.1146	1.31(0.56-3.05)	0.54
	C5	39	128	23.35		2.13(1.43-3.17)	0.0002	1.35(0.62-2.92)	0.45
Cirrhosis	C1	74	434	14.57	13.02	1	-	1	-
	C2	47	414	10.20		0.69(0.48-1.00)	0.0511	0.65(0.48-0.89)	0.0074
	C3	50	405	10.99		0.82(0.57-1.17)	0.2659	0.65(0.47-0.89)	0.0074
	C4	26	114	18.57		1.46(0.94-2.29)	0.0942	1.08(0.72-1.60)	0.71
	C5	28	136	17.07		1.18(0.77-1.83)	0.4446	0.95(0.64-1.40)	0.78
HCC	C1	30	534	5.32	3.95	1	-	1	-
	C2	20	465	4.12		0.77(0.44-1.35)	0.3568	0.69(0.41-1.17)	0.17
	C3	8	463	1.70		0.35(0.16-0.76)	0.0078	0.28(0.13-0.61)	0.0012
	C4	7	149	4.49		0.91(0.40-2.07)	0.8214	0.94(0.46-1.94)	0.87
	C5	8	162	4.71		0.89(0.41-1.95)	0.7756	0.82(0.40-1.69)	0.59
ARF	C1	25	558	4.29	4.5	1	-	-	-
	C2	25	461	5.14		1.08(0.62-1.87)	0.7973	-	-
	C3	15	464	3.13		0.63(0.33-1.20)	0.1642	-	-
	C4	2	166	1.19		0.29(0.07-1.23)	0.0923	-	-
	C5	18	155	10.40		2.10(1.14-3.85)	0.0165	-	-
IHD	C1	22	514	4.10	4.61	1	-	-	-
	C2	29	421	6.44		1.47(0.84-2.56)	0.1735	-	-
	C3	22	439	4.77		1.07(0.59-1.93)	0.8231	-	-
	C4	0	168	0.00		0.00(0.00-Inf)	0.9928	-	-
	C5	9	153	5.56		1.15(0.53-2.50)	0.7223	-	-
Major depression	C1	35	515	6.36	8.5	1	-	-	-
	C2	35	452	7.19		1.10(0.69-1.76)	0.6847	-	-
	C3	52	348	13.00		1.88(1.23-2.89)	0.0038	-	-
	C4	10	144	6.49		1.06(0.52-2.13)	0.8813	-	-
	C5	18	155	10.40		1.58(0.90-2.80)	0.1123	-	-
Sleep Apnea	C1	34	521	6.13	11.38	1	-	-	-
	C2	67	355	15.88		2.58(1.71-3.89)	6.98×10^{-6}	-	-
	C3	67	355	15.88		2.49(1.64-3.75)	1.54×10^{-5}	-	-
	C4	8	153	4.97		0.82(0.38-1.77)	0.6086	-	-
	C5	20	143	12.27		1.84(1.06-3.20)	0.0299	-	-

The LCA-identified subtypes in the development cohort and the centroid-assigned subtypes in the validation cohort were combined. Subtype C1 was used as the reference group in Cox regression analysis and Fine-Gray competing risk analysis. HR (95% CI), Hazard Ratio with 95% Confidence Interval. Two-sided P values were obtained from Wald tests. HCC, hepatocellular carcinoma; MASH, metabolic dysfunction-associated steatohepatitis; IHD, ischemic heart disease; ARF, acute renal failure.

Supplementary Table 9: Changes of biomarkers over six years within each subgroup in the MCB cohort

Variable	Subgroup	No. patients	Mean of year 2017	Mean of year 2023	Change (%)	P-value	P-adjusted	P-value (format)	P-adjusted (format)	P-significance
BMI	C1	275	28.4	27.1	-4.6	6.66E-06	3.33E-05	<0.0001	<0.0001	****
	C2	217	36	34.5	-4.2	2.75E-11	1.38E-10	<0.0001	<0.0001	****
	C3	240	37.1	35	-5.7	1.45E-11	7.25E-11	<0.0001	<0.0001	****
	C4	59	31.3	31.2	-0.3	0.876	1	0.876	1	ns
	C5	63	31.2	30.3	-2.9	0.0654	0.327	0.065	0.327	ns
ALT	C1	146	41.2	30.7	-25.5	0.00041	0.00205	<0.001	0.002	***
	C2	103	42.1	32.9	-21.9	3.35E-05	0.00017	<0.0001	<0.001	****
	C3	127	35.5	25.9	-27	1.11E-06	5.55E-06	<0.0001	<0.0001	****
	C4	45	39.2	30.6	-21.9	0.166	0.83	0.166	0.83	ns
	C5	44	59.5	62.5	5	0.55	1	0.55	1	ns
AST	C1	136	33.4	28.6	-14.4	0.0115	0.0575	0.011	0.057	*
	C2	66	33	32.2	-2.4	0.771	1	0.771	1	ns
	C3	96	36.2	26.5	-26.8	0.0283	0.1415	0.028	0.141	*
	C4	28	31.1	27.8	-10.6	0.309	1	0.309	1	ns
	C5	28	48.2	63.9	32.6	0.033	0.165	0.033	0.165	*
ALP	C1	176	95.4	94.2	-1.3	0.823	1	0.823	1	ns
	C2	118	85.8	97.1	13.2	0.00265	0.01325	0.003	0.013	**
	C3	142	96.6	101.6	5.2	0.29	1	0.29	1	ns
	C4	51	93.4	96.5	3.3	0.687	1	0.687	1	ns
	C5	46	103.5	155.8	50.5	0.00018	0.00088	<0.001	<0.001	***
LDL	C1	234	94.2	83.9	-10.9	0.00097	0.00483	<0.001	0.005	***
	C2	224	83.5	69.7	-16.5	7.35E-07	3.68E-06	<0.0001	<0.0001	****
	C3	218	100.2	89.5	-10.7	0.00161	0.00805	0.002	0.008	**
	C4	42	86.6	88.7	2.4	0.728	1	0.728	1	ns
	C5	51	96	85.1	-11.4	0.12	0.6	0.12	0.6	ns
HDL	C1	99	50.6	52	2.8	0.558	1	0.558	1	ns
	C2	147	41.1	41.4	0.7	0.779	1	0.779	1	ns
	C3	138	51.7	53.7	3.9	0.278	1	0.278	1	ns
	C4	30	59.1	59.7	1	0.904	1	0.904	1	ns
	C5	31	49.6	50.4	1.6	0.839	1	0.839	1	ns
TG	C1	244	156	139.6	-10.5	0.023	0.115	0.023	0.115	*
	C2	237	187	157.3	-15.9	0.00243	0.01215	0.002	0.012	**
	C3	227	181.9	164.9	-9.3	0.032	0.16	0.032	0.16	*
	C4	46	106.4	102.1	-4	0.658	1	0.658	1	ns
	C5	54	167.1	156.9	-6.1	0.614	1	0.614	1	ns
HbA1c	C1	212	6.6	6.6	0	0.936	1	0.936	1	ns
	C2	253	7	6.8	-2.9	0.138	0.69	0.138	0.69	ns
	C3	226	7	6.8	-2.9	0.157	0.785	0.157	0.785	ns
	C4	30	5.4	5.5	1.9	0.341	1	0.341	1	ns
	C5	44	6.7	7	4.5	0.28	1	0.28	1	ns

The LCA-identified subgroups in the development cohort and the centroid-assigned subgroups in the validation cohort were combined. Change (%) = (Mean of year 2023 - Mean of year 2017)*100/Mean of year 2017. Significant test used two-sided paired t-test with Bonferroni adjustment. Statistical significance: ****P<0.0001; ***P<0.001; **P<0.01; *P<0.05; ns, not significant. ALT, alanine aminotransferase; AST, aspartate aminotransferase; ALP, alkaline phosphatase; BMI, body mass index; HbA1c, hemoglobin A1C; HDL, high-density lipoprotein; LDL, low-density lipoprotein; TG, triglycerides.

Supplementary Table 10: Risk of liver transplantation in the two cohorts

Cohort	Membership Method	Subgroup	No. of events	Censored	% events	HR (95% CI)	P value
MCB	Baseline + Centroid	C1	94	465	16.82	1	-
		C2	46	442	9.43	0.55(0.39-0.78)	0.0009
		C3	24	447	5.10	0.32(0.21-0.50)	6.77E-07
		C4	25	130	16.13	1.04(0.67-1.62)	0.8482
		C5	17	156	9.83	0.57(0.34-0.95)	0.0327
Tapestry	Probability	C1	70	658	9.62	1	-
		C2	17	228	6.94	0.71(0.42-1.21)	0.209
		C3	24	706	3.29	0.35(0.22-0.55)	7.99E-06
		C4	44	397	9.98	1.07(0.74-1.57)	0.7084
		C5	21	199	9.55	1.12(0.69-1.82)	0.6538
	Centroid	C1	60	572	9.49	1	-
		C2	32	434	6.87	0.69(0.45-1.06)	0.0913
		C3	21	613	3.31	0.34(0.21-0.57)	2.60E-05
		C4	45	419	9.70	1.04(0.70-1.52)	0.8606
		C5	18	150	10.71	1.25(0.74-2.11)	0.4088
	Core Points	C1	57	557	9.28	1	-
		C2	33	435	7.05	0.73(0.47-1.11)	0.1434
		C3	21	602	3.37	0.36(0.22-0.59)	0.0001
		C4	47	444	9.57	1.06(0.72-1.55)	0.7806
		C5	18	150	10.71	1.26(0.74-2.14)	0.3918

For MCB, subgroups identified by LCA in the development cohort ("Baseline") and subgroups assigned by centroid in the validation cohort were combined. Subgroup C1 was used as the reference group in Cox regression analysis. HR (95% CI), Hazard Ratio (95% Confidence Interval). Two-sided P values were obtained from Wald tests.