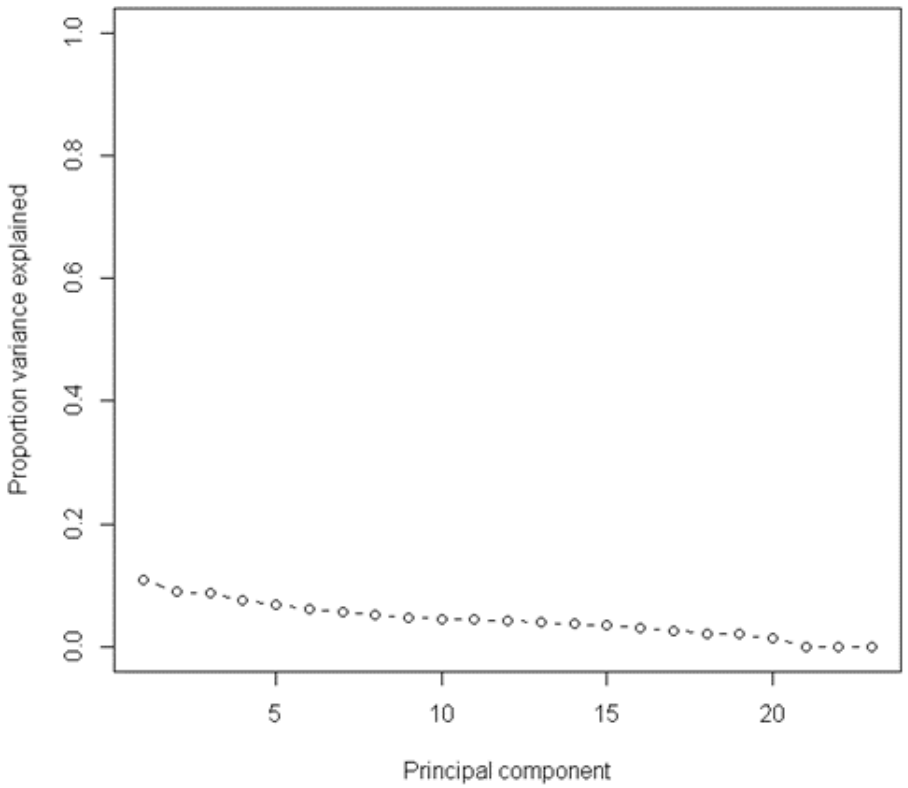
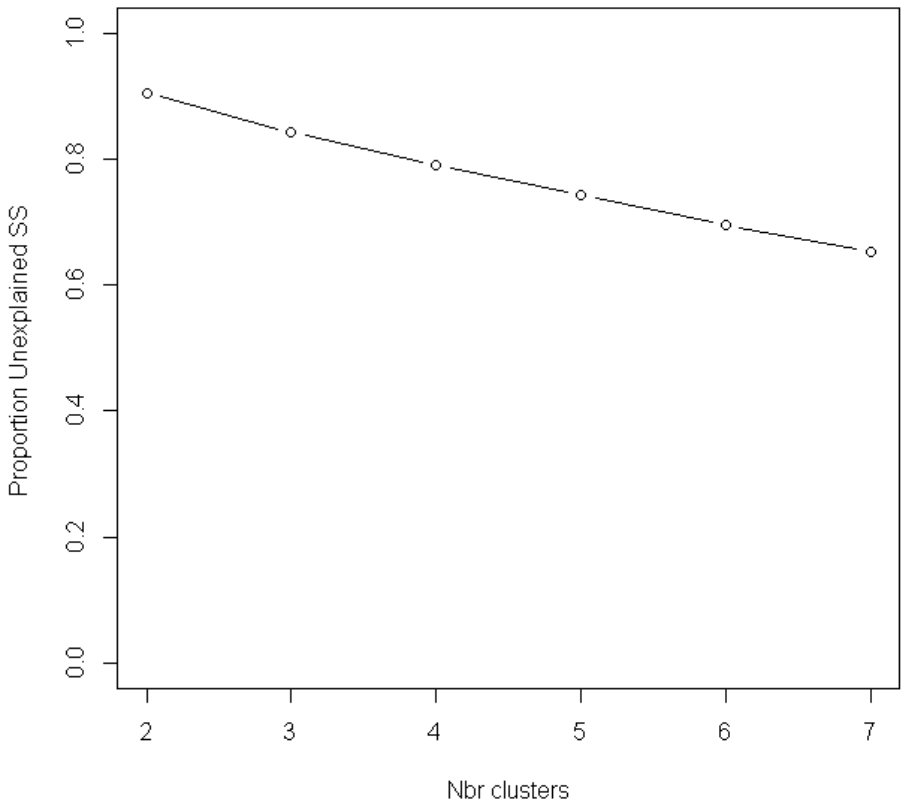
Supplementary Table 1. SNPs underlying the pathway-based genetic risk scores for age-related macular degeneration.

Suggested Pathway of AMD Locus	dbSNP ID	Chr	Pos. (hg19)	RefSeq Gene Closest to SNP	Risk Allele	Other Allele	Weight / log(OR)
Complement Pathways	rs187328863	1	196380158	*KCNT2*	T	C	0.38623
Complement Pathways	rs148553336	1	196613173	*CFH*	T	C	1.1637
Complement Pathways	rs570618	1	196657064	*CFH*	T	G	0.55141
Complement Pathways	rs10922109	1	196704632	*CFH*	C	A	0.66958
Complement Pathways	rs35292876	1	196706642	*CFH*	T	C	0.43398
Complement Pathways	rs121913059	1	196716375	*CFH*	T	C	3.8635
Complement Pathways	rs61818925	1	196815450	*CFHR1*	T	G	0.16257
Complement Pathways	rs191281603	1	196958651	*CFHR5*	C	G	0.89702
Extracellular Matrix Remodeling	rs11884770	2	228086920	*LOC654841*	C	T	0.080248
Extracellular Matrix Remodeling	rs62247658	3	64715155	*ADAMTS9*	C	T	0.12671
Extracellular Matrix Remodeling	rs140647181	3	99180668	*COL8A1*	C	T	0.61508
Extracellular Matrix Remodeling	rs55975637	3	99419853	*COL8A1*	A	G	0.14899
Complement Pathways	rs10033900	4	110659067	*CFI*	T	C	0.1427
Complement Pathways	rs141853578	4	110685820	*CFI*	T	C	1.633
…	rs114092250	5	35494448	*SPEF2*	G	A	0.33716
Complement Pathways	rs62358361	5	39327888	*C9*	T	G	0.51329
Complement Pathways	rs116503776	6	31930462	*SKIV2L*	G	A	0.66418
Complement Pathways	rs144629244	6	31946792	*STK19*	A	G	1.0253
Complement Pathways	rs181705462	6	31947027	*STK19*	T	G	0.44737
Complement Pathways	rs114254831	6	32155581	*PBX2*	G	A	0.12483
Extracellular Matrix Remodeling	rs943080	6	43826627	*LINC01512*	T	C	0.13934
…	rs7803454	7	99991548	*PILRA*	T	C	0.14304
…	rs1142	7	104756326	*SRPK2*	T	C	0.13057
…	rs79037040	8	23082971	*LOC389641*	T	G	0.11331
…	rs71507014	9	73438605	*TRPM3*	G	GC	0.11024
…	rs10781182	9	76617720	*LOC101927358*	T	G	0.10031
…	rs1626340	9	101923372	*TGFBR1*	G	A	0.12318
Lipid Metabolism	rs2740488	9	107661742	*ABCA1*	A	C	0.11123
…	rs12357257	10	24999593	*ARHGAP21*	A	G	0.11008
ARMS2/HTRA1	rs3750846	10	124215565	*ARMS2*	C	T	1.0744
…	rs3138141	12	56115778	*FAM138D*	A	C	0.16536
…	rs61941274	12	112132610	*ACAD10*	A	G	0.46924
…	rs9564692	13	31821240	*B3GLCT*	C	T	0.10425
…	rs61985136	14	68769199	*RAD51B*	T	C	0.12523
…	rs2842339	14	68986999	*OR11H12*	G	A	0.16483
Lipid Metabolism	rs2043085	15	58680954	*LIPC*	C	T	0.14337
Lipid Metabolism	rs2070895	15	58723939	*LIPC*	G	A	0.15207
Lipid Metabolism	rs17231506	16	56994528	*CETP*	T	C	0.10813
Lipid Metabolism	rs5817082	16	56997349	*CETP*	C	CA	0.13584
…	rs72802342	16	75234872	*CTRB2*	C	A	0.22983
Complement Pathways	rs11080055	17	26649724	*TMEM97*	C	A	0.083858
…	rs6565597	17	79526821	*NPLOC4*	T	C	0.10928
…	rs67538026	19	1031438	*CNN2*	C	T	0.1053
Complement Pathways	rs12019136	19	5835677	*FUT6*	G	A	0.29875
Complement Pathways	rs147859257	19	6718146	*C3*	G	T	1.1683
Complement Pathways	rs2230199	19	6718387	*C3*	G	C	0.38789
Lipid Metabolism	rs429358	19	45411941	*APOE*	T	C	0.39882
Lipid Metabolism	rs73036519	19	45748362	*MARK4*	G	C	0.092776
…	rs142450006	20	44614991	*ZNF335*	TTTTC	T	0.17438
…	rs201459901	20	56653724	*C20orf85*	T	TA	0.28024
Extracellular Matrix Remodeling	rs5754227	22	33105817	*TIMP3*	T	C	0.23796
…	rs8135665	22	38476276	*SLC16A8*	T	C	0.13406

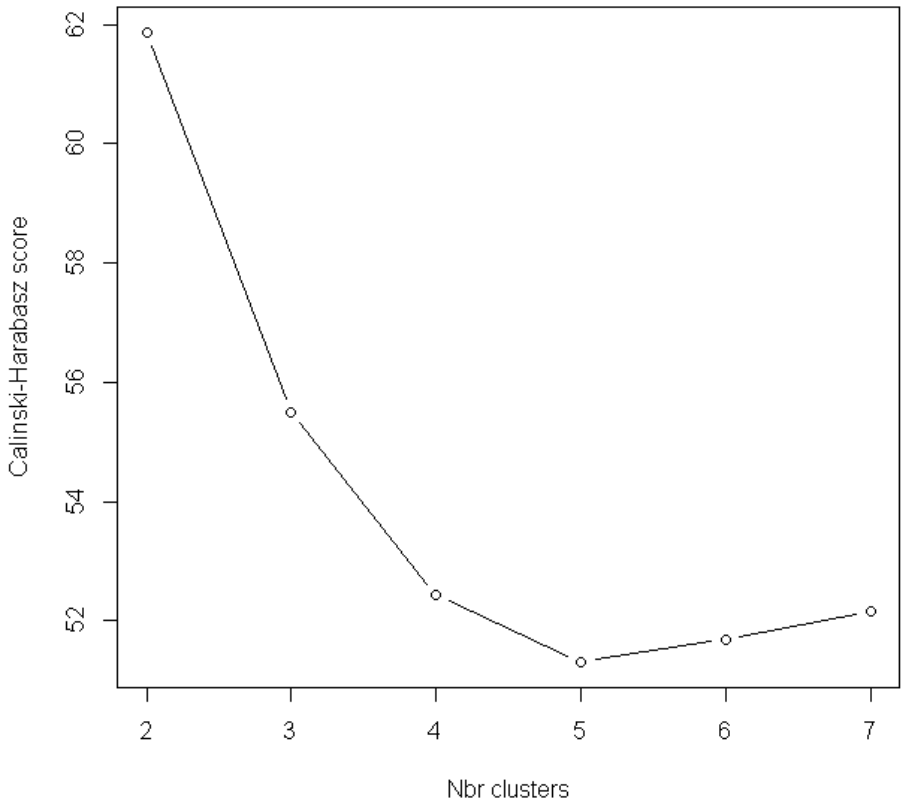
Supplementary Figure 1. Scree plot showing the proportion of variance explained by each principal component, following principal components analysis applied to the cross-sectional dataset of phenotypic characteristics.



Supplementary Figure 2. Plot showing the proportion of the unexplained sums of squares according to cluster number, following k-means cluster analysis applied to the cross-sectional dataset of phenotypic characteristics.

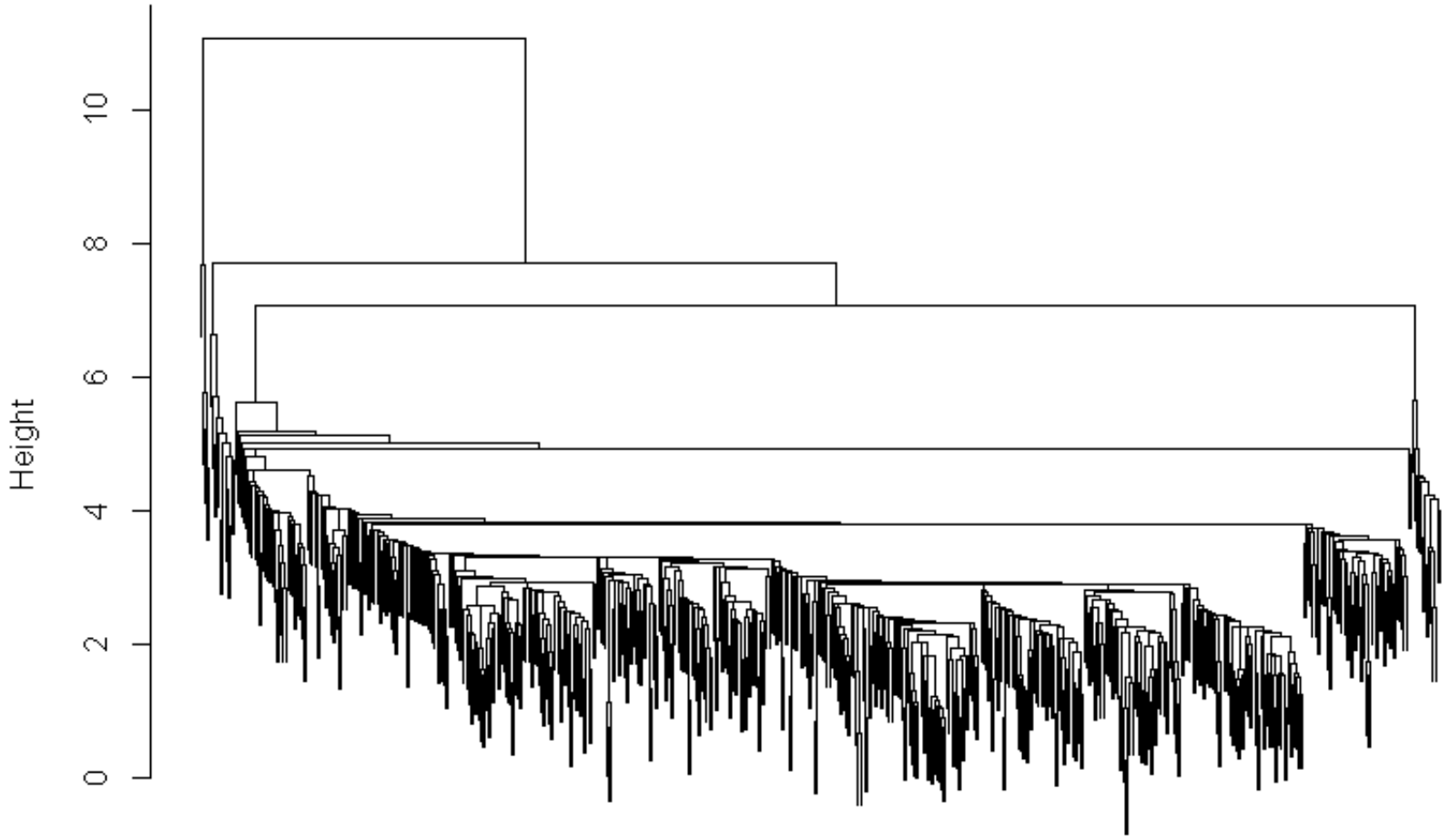


Supplementary Figure 3. Plot showing the Calinski-Harabasz scores according to cluster number, following k-means cluster analysis applied to the cross-sectional dataset of phenotypic characteristics.

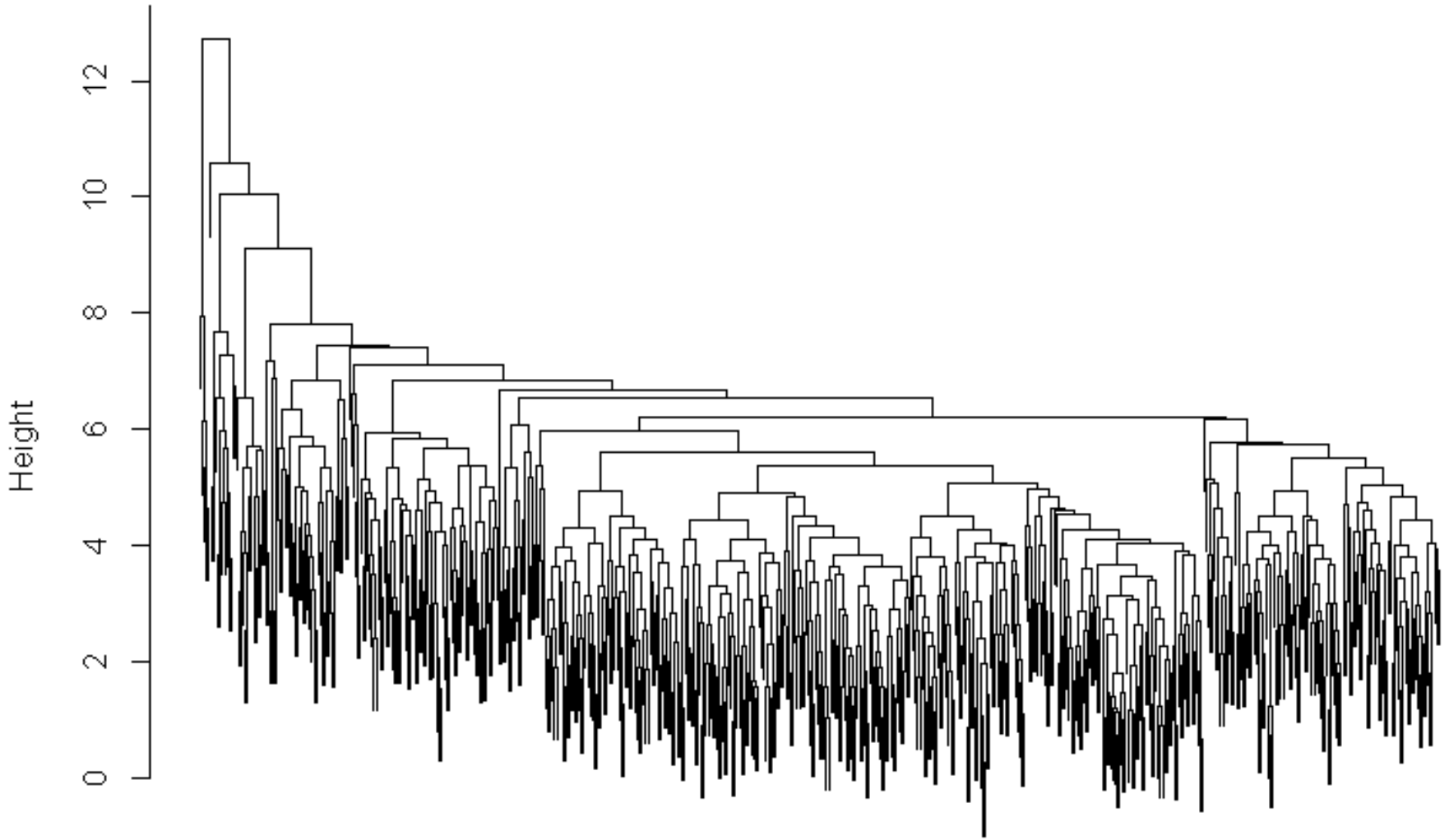


Supplementary Figure 4. Dendrograms following agglomerative hierarchical cluster analysis applied to the cross-sectional dataset of phenotypic characteristics, according to linkage type. Dissimilarity is plotted on the y-axis and each participant is shown on the x-axis. Each horizontal line represents the fusion of a pair of clusters, with the height of the segment showing the dissimilarity between the members of the pair. Clusters that fuse near the bottom of the tree are more similar, while clusters that fuse near the top are less similar.

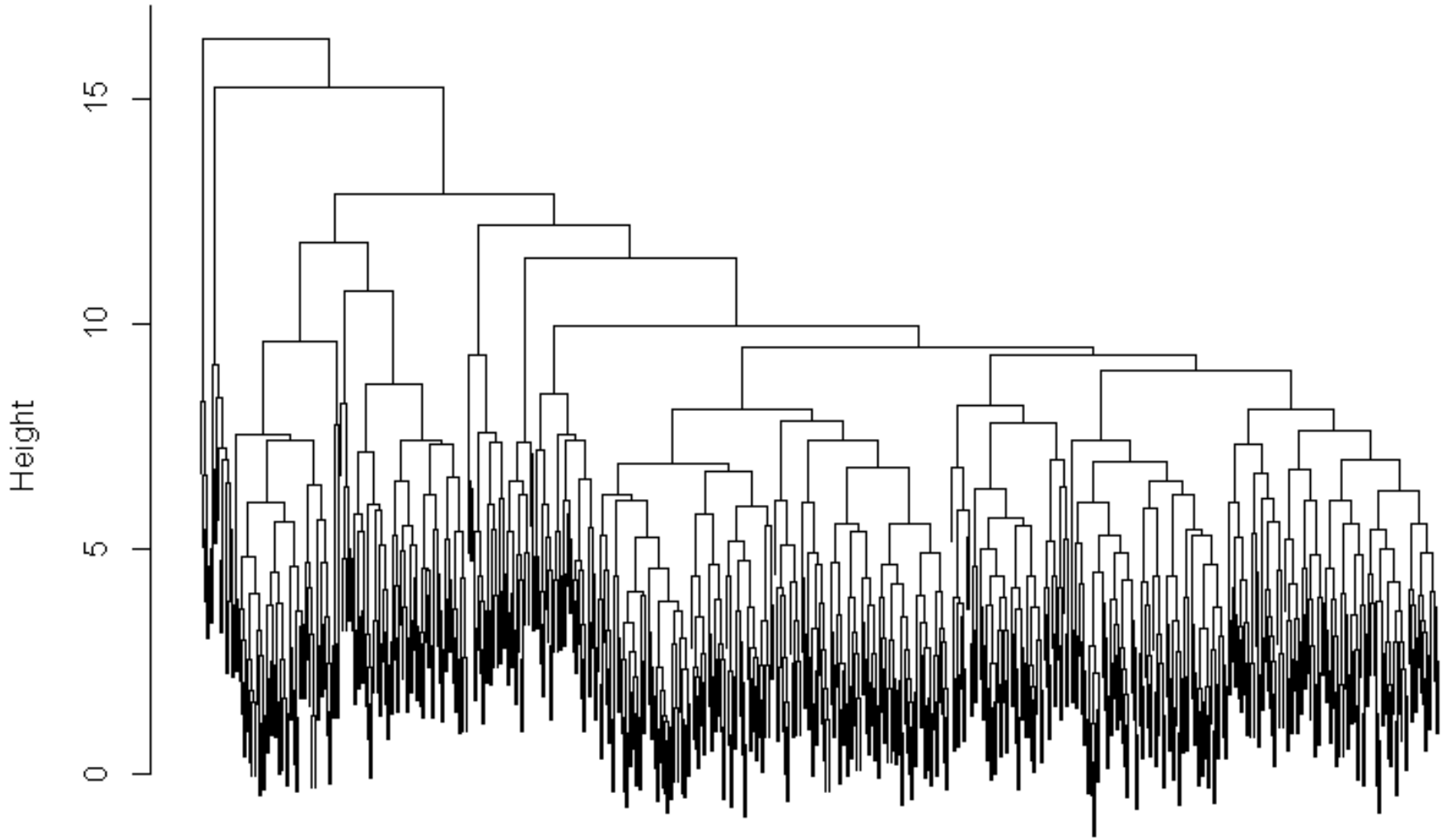
Single linkage (cophenetic correlation coefficient 0.83)



Average linkage (cophenetic correlation coefficient 0.88)

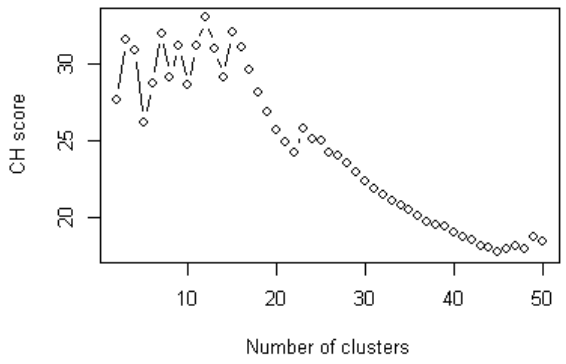


Complete linkage (cophenetic correlation coefficient 0.63)

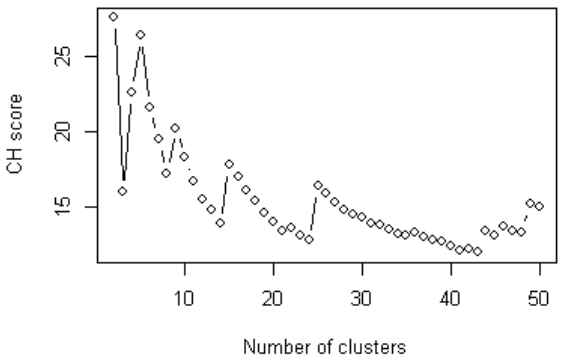


Supplementary Figure 5. Plot showing the Calinski-Harabasz scores according to cluster number, following agglomerative hierarchical cluster analysis applied to the cross-sectional dataset of phenotypic characteristics, according to linkage type.

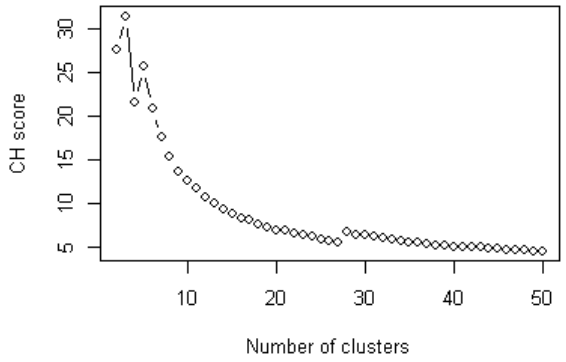
Complete linkage



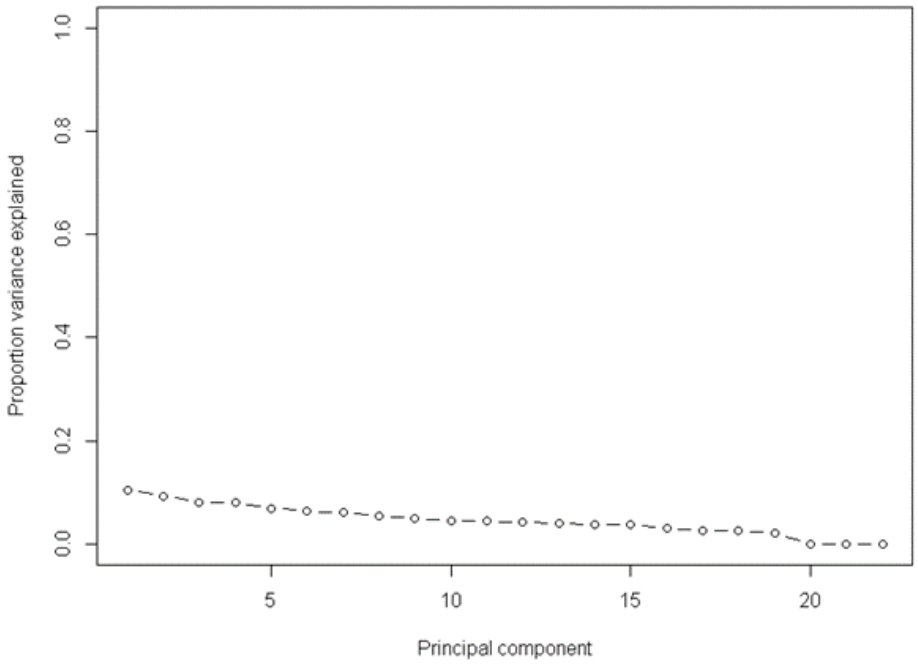
Average linkage



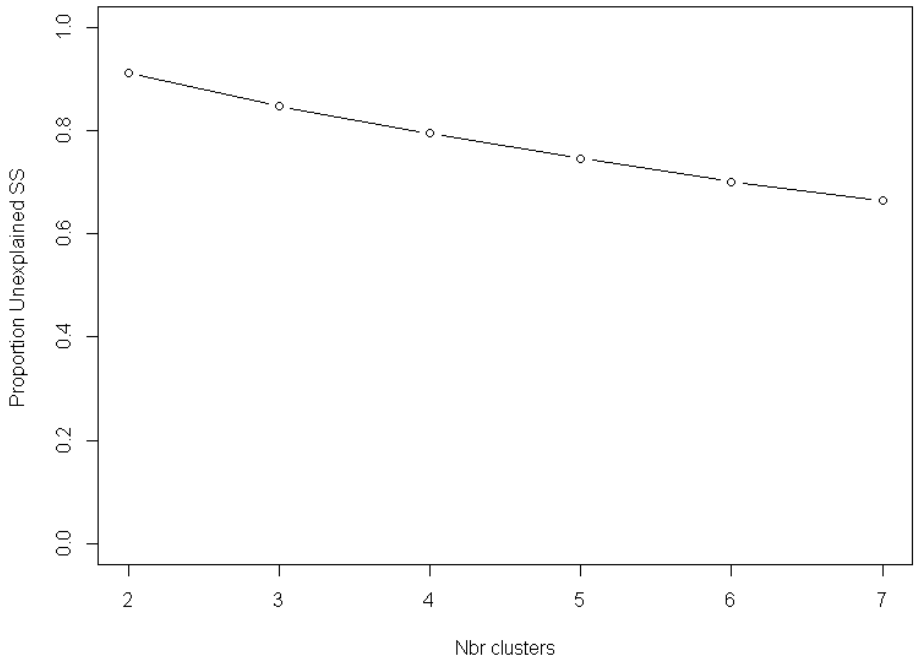
Single linkage



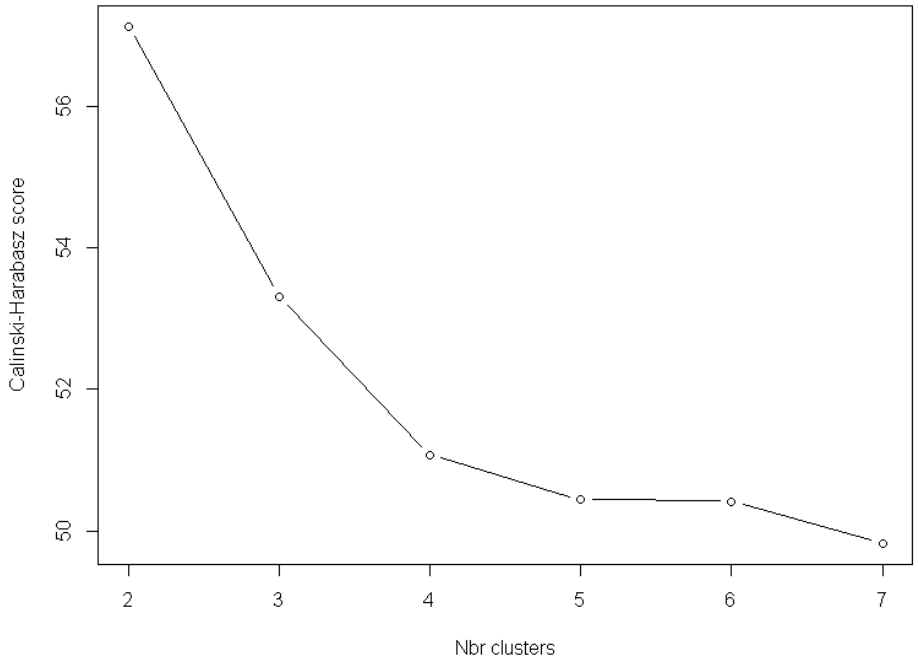
Supplementary Figure 6. Scree plot showing the proportion of variance explained by each principal component, following principal components analysis applied to the longitudinal dataset of phenotypic characteristics.



Supplementary Figure 7. Plot showing the proportion of the unexplained sums of squares according to cluster number, following k-means cluster analysis applied to the longitudinal dataset of phenotypic characteristics.

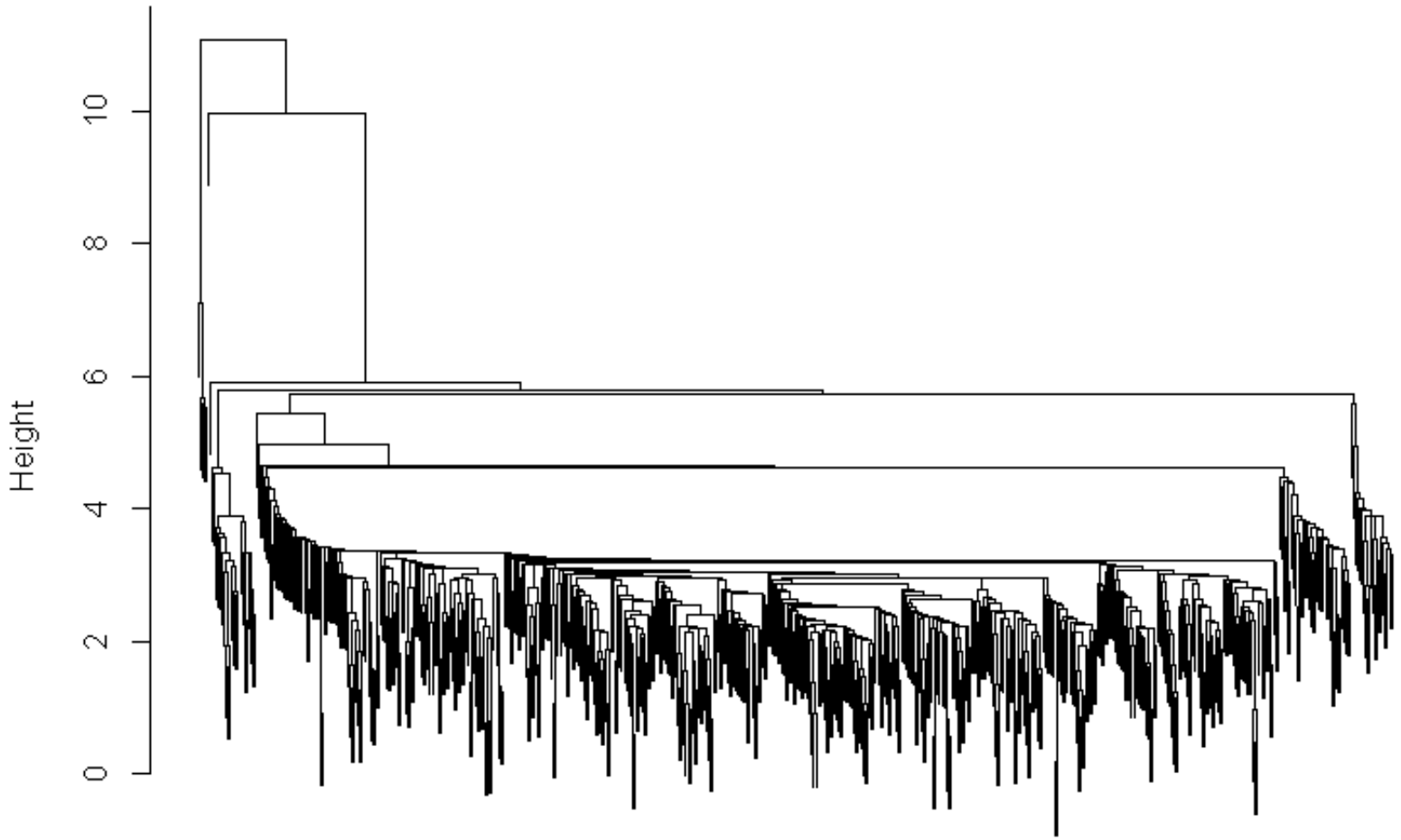


Supplementary Figure 8. Plot showing the Calinski-Harabasz scores according to cluster number, following k-means cluster analysis applied to the longitudinal dataset of phenotypic characteristics.

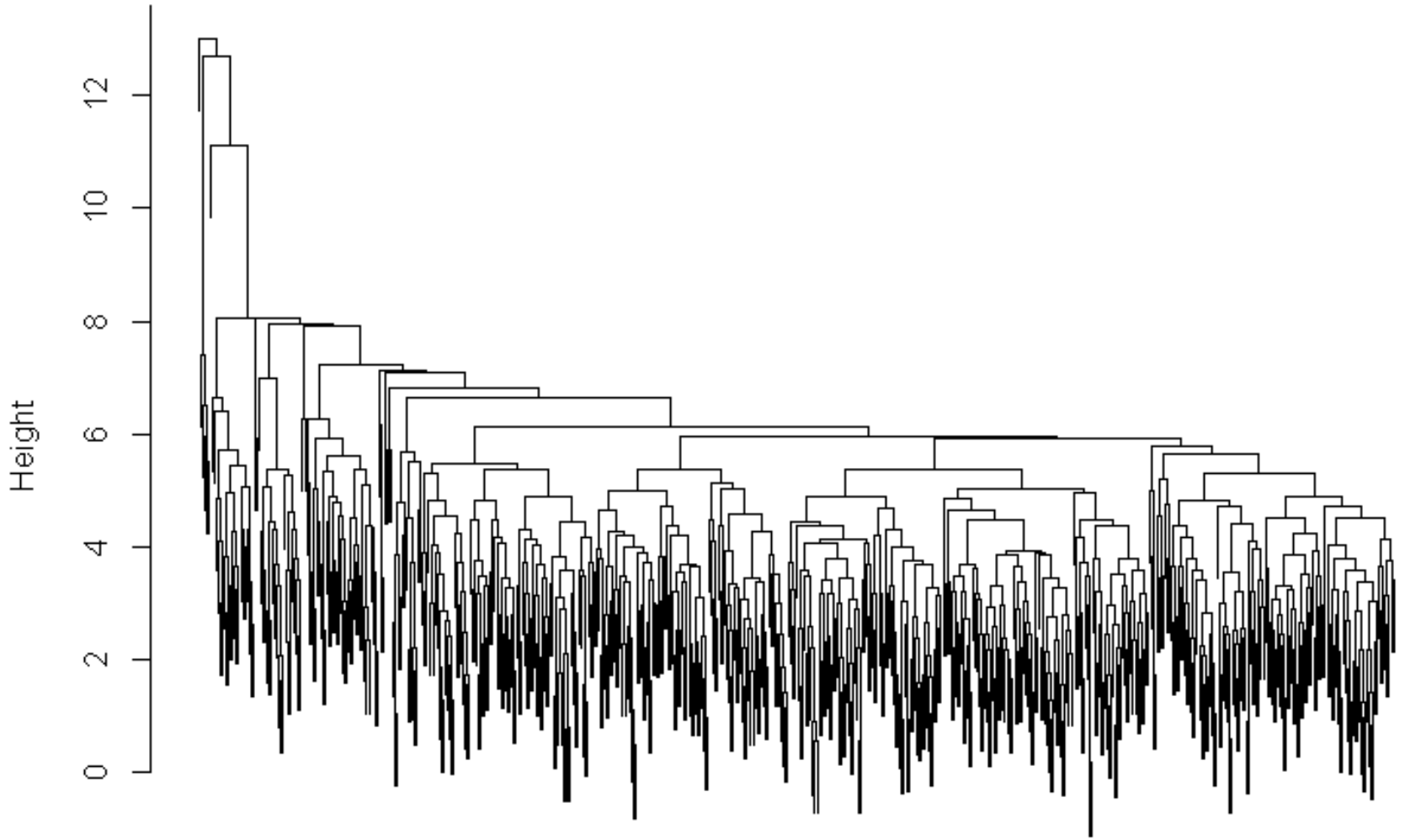


Supplementary Figure 9. Dendrograms following agglomerative hierarchical cluster analysis applied to the longitudinal dataset of phenotypic characteristics, according to linkage type. Dissimilarity is plotted on the y-axis and each participant is shown on the x-axis. Each horizontal line represents the fusion of a pair of clusters, with the height of the segment showing the dissimilarity between the members of the pair. Clusters that fuse near the bottom of the tree are more similar, while clusters that fuse near the top are less similar.

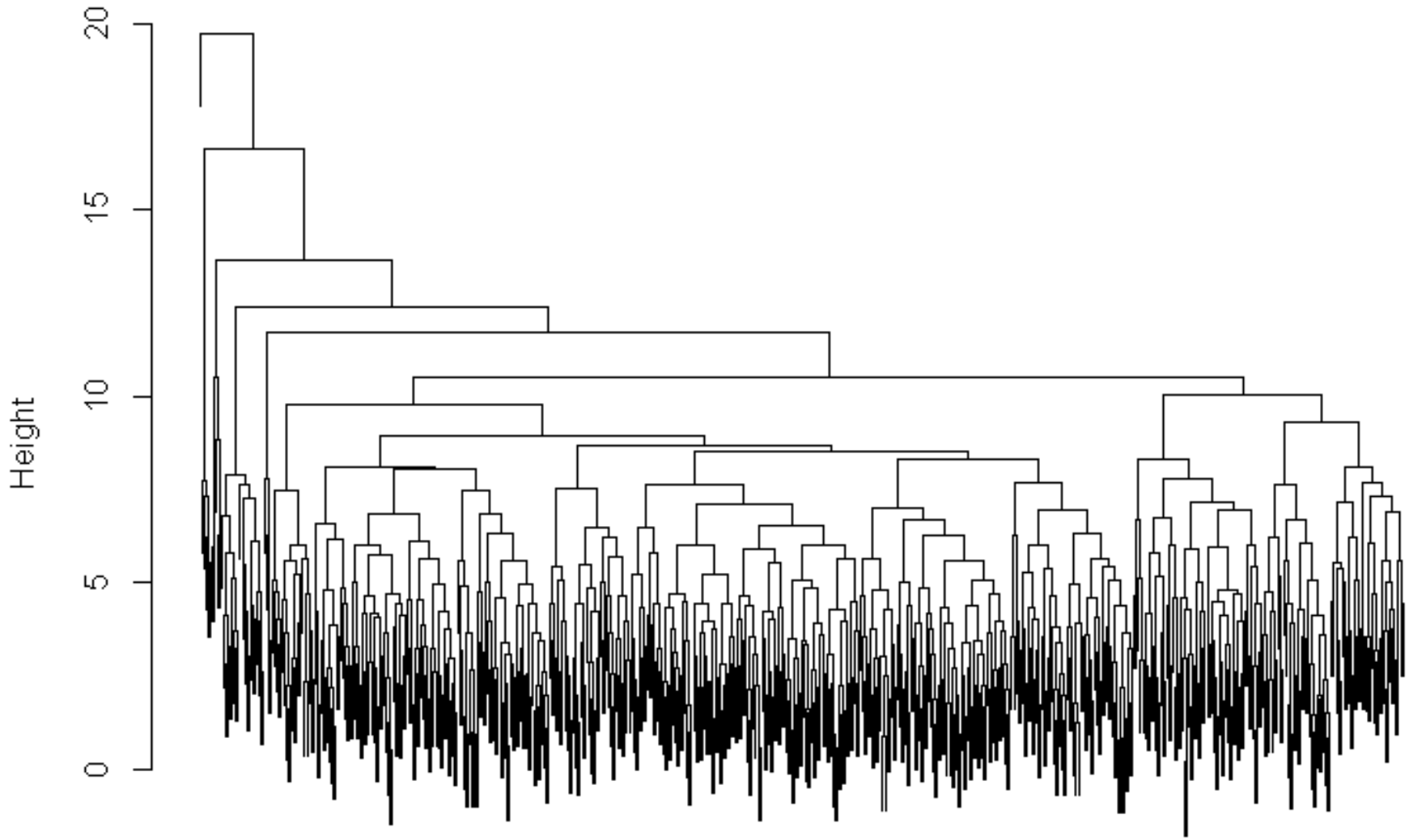
Single linkage (cophenetic correlation coefficient 0.82)



Average linkage (cophenetic correlation coefficient 0.88)

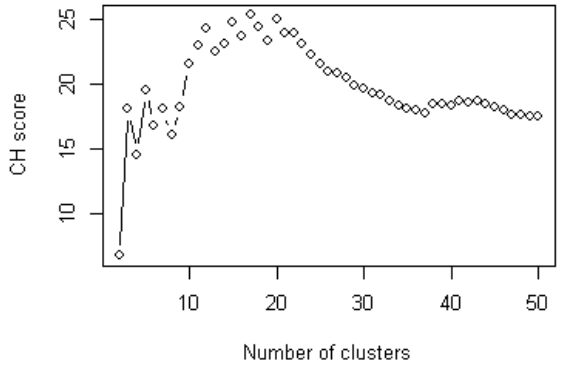


Complete linkage (cophenetic correlation coefficient 0.73)

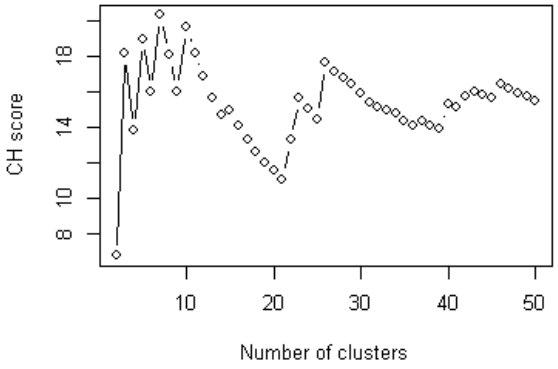


Supplementary Figure 10. Plot showing the Calinski-Harabasz scores according to cluster number, following agglomerative hierarchical cluster analysis applied to the longitudinal dataset of phenotypic characteristics, according to linkage type.

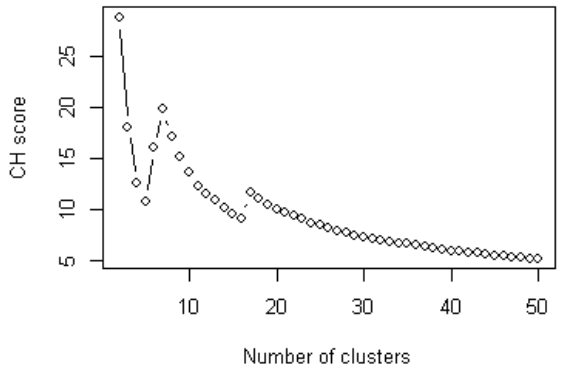
Complete linkage



Average linkage



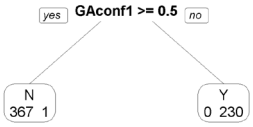
Single linkage



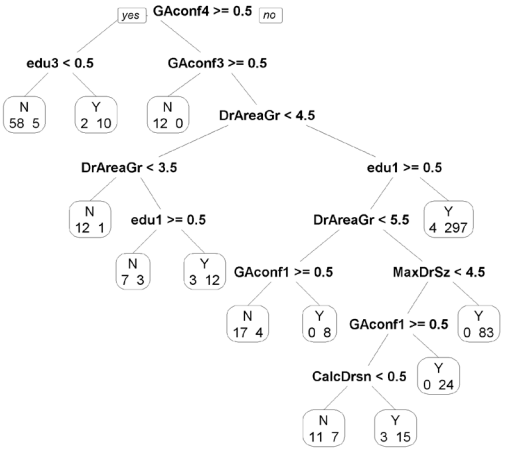
Supplementary Figure 11. CART classification trees and related confusion matrices and performance metrics for the clusters identified by phenotypic characteristics, based on CART classification by the same phenotypic characteristics.

A

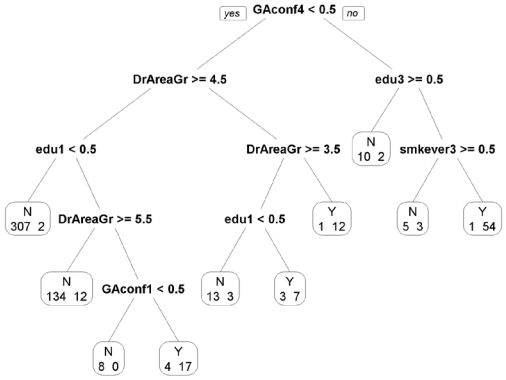
B



C



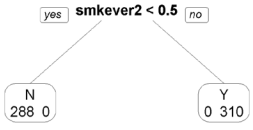
D



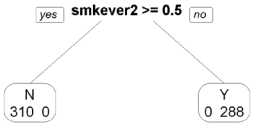
E

F*

G



H



* Classification by CART was not possible for cluster F, owing to very small numbers (n=5).

Cluster A	Cluster B	Cluster C
Confusion matrix	Confusion matrix	Confusion matrix
predicted N Y	predicted N Y	predicted N Y
N 230 0	N 367 1	N 117 20
Y 1 367	Y 0 230	Y 12 449
Accuracy = 0.998328	Accuracy = 0.998328	Accuracy = 0.946488
Sensitivity = 1.000000	Sensitivity = 0.995671	Sensitivity = 0.957356
Specificity = 0.995671	Specificity = 1.000000	Specificity = 0.906977

Cluster D	Cluster E	
Confusion matrix	Confusion matrix	
predicted N Y	predicted N Y	
N 477 22	N 586 0	
Y 9 90	Y 0 12	
Accuracy = 0.948161	Accuracy = 1.000000	
Sensitivity = 0.803571	Sensitivity = 1.000000	
Specificity = 0.981481	Specificity = 1.000000	

Cluster G	Cluster H	
Confusion matrix	Confusion matrix	
predicted N Y	predicted N Y	
N 597 1	N 584 0	
Y 0 0	Y 0 14	
Accuracy = 0.998328	Accuracy = 1.000000	
Sensitivity = 0.000000	Sensitivity = 1.000000	
Specificity = 1.000000	Specificity = 1.000000	

Supplementary Figure 12. Results of logistic regression with LASSO, with related confusion matrices and performance metrics, for the clusters identified by phenotypic characteristics, based on logistic regression according to the same phenotypic characteristics.

Log odds ratio estimates

NAME	A	B	C	D	E
DrAreaGrd	.	.	0.34	-0.51	.
MaxDrSz	.	.	0.17	-0.27	.
edu1	.	.	-0.08	0.44	.
edu3	.	.	.	-0.04	.
GAconf1	1.02	-1.02	.	.	.
GAconf3	.	.	-1.82	.	5.04
GAconf4	.	.	-2.19	2.70	.
GAconf5	.	.	.	.	.

Variables never selected: white, age, male, GAconf2, GAsqrt, GArate, CGA, CalcDrsn, GA_FE, RPDprob, VA, edu2, smkever1, smkever2, and smkever3.

* Logistic regression not possible for cluster F, owing to very small numbers (n=5).

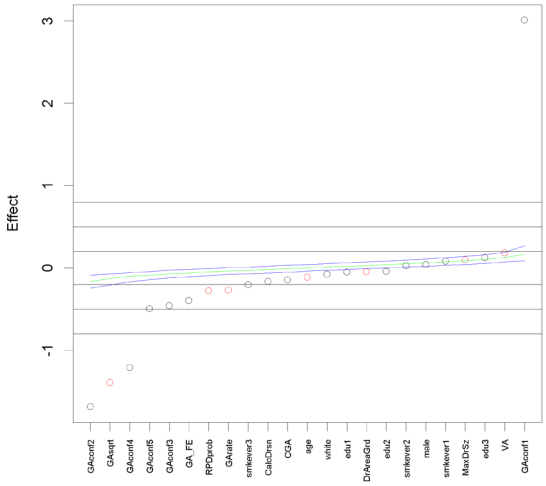
Name	G	H
GAconf3	.	.
smkever2	0.36	-0.36

Variables never selected: age, male, white, GArate, CGA, CalcDrsn, DrAreaGrd, MaxDrSz, GA_FE, RPDprob, VArate, edu1, edu2, edu3, GAconf1, GAconf2, GAconf4, GAconf5, smkever1, and smkever3.

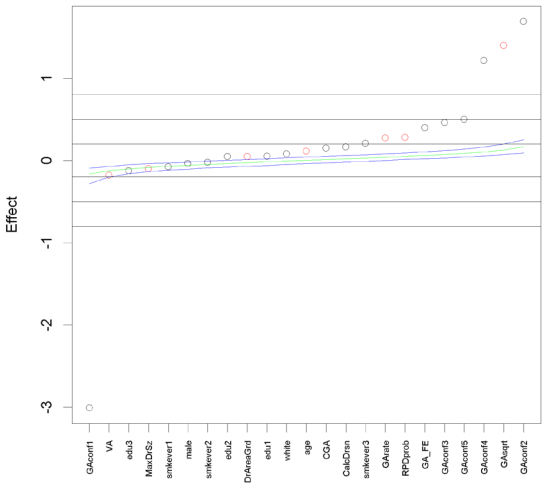
```
Cluster A              | Cluster B              | Cluster C
Confusion matrix       | Confusion matrix       | Confusion matrix
pred   N   Y           | pred   N   Y           | pred   N   Y
   N 230   0           |    N 367   1           |    N  57  11
   Y   1 367           |    Y   0 230           |    Y  72 458
Accuracy = 0.998328    | Accuracy = 0.998328    | Accuracy = 0.861204
Sensitivity = 1.000000 | Sensitivity = 0.995671 | Sensitivity = 0.976546
Specificity = 0.995671 | Specificity = 1.000000 | Specificity = 0.441860
-----------------------+|-----------------------+------------------------
Cluster D              | Cluster E              |
Confusion matrix       | Confusion matrix       |
pred   N   Y           | pred   N   Y           |
   N 474  57           |    N 586   0           |
   Y  12  55           |    Y   0  12           |
Accuracy = 0.884615    | Accuracy = 1.000000    |
Sensitivity = 0.491071 | Sensitivity = 1.000000 |
Specificity = 0.975309 | Specificity = 1.000000 |
-----------------------+|-----------------------+------------------------
Cluster G              | Cluster H              |
Confusion matrix       | Confusion matrix       |
    y                  |     y                  |
pred   N   Y           | pred   N   Y           |
   N 288   0           |    N 310   0           |
   Y   0 310           |    Y   0 288           |
Accuracy = 1.000000    | Accuracy = 1.000000    |
Sensitivity = 1.000000 | Sensitivity = 1.000000 |
Specificity = 1.000000 | Specificity = 1.000000 |
```

Supplementary Figure 13. Cohen's effect sizes for the clusters identified by phenotypic characteristics, based on effect sizes for the same phenotypic characteristics.

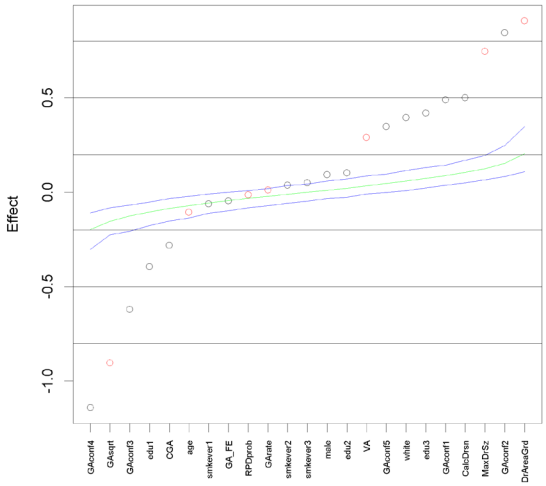
Cluster A



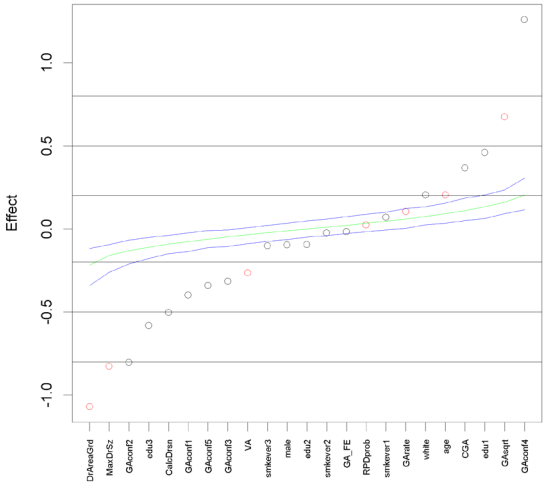
Cluster B



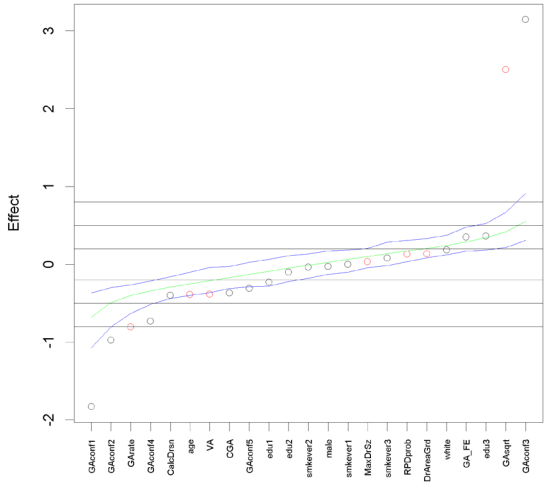
Cluster C



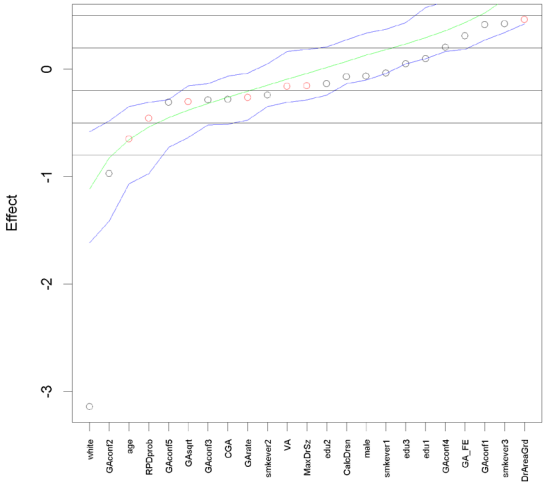
Cluster D



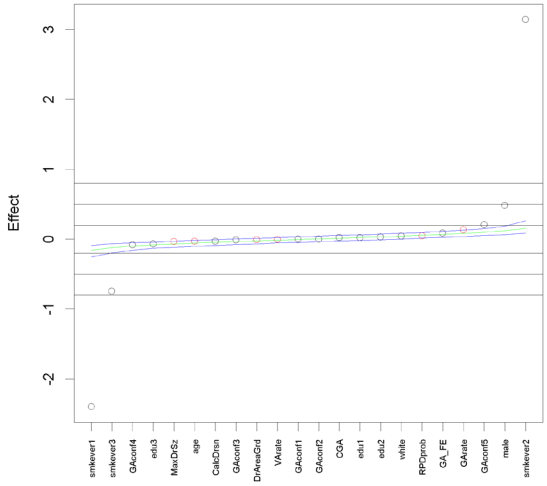
Cluster E



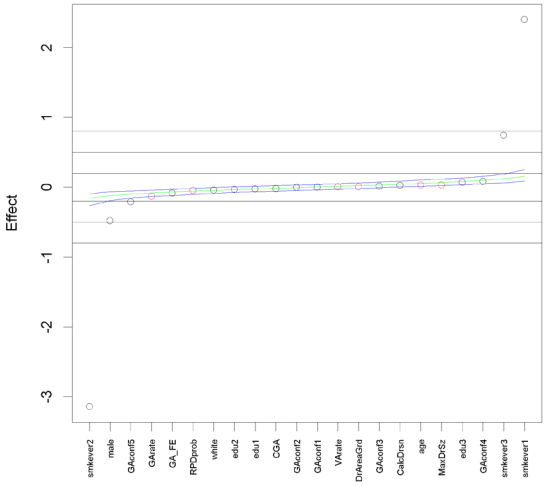
Cluster F



Cluster G



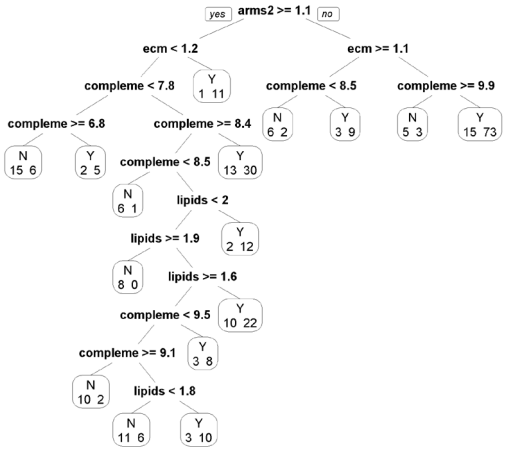
Cluster H



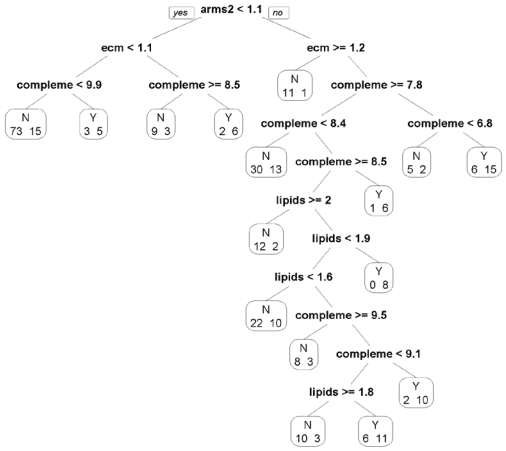
The effect sizes are plotted in ascending order on the y axis. Horizontal reference lines mark effect sizes for +/-0.2, +/-0.5, and +/-0.8. Points are colored red or black according to whether they are for Cohen's d (continuous features) or Cohen's h (binary features). The green and blue curves are expected values and 95% confidence intervals from a simulation of 1000 iterations under the null hypothesis that the cluster is unrelated to the features.

Supplementary Figure 14. CART classification trees and related confusion matrices and performance metrics for the clusters identified by phenotypic characteristics, based on CART classification by the genetic characteristics.

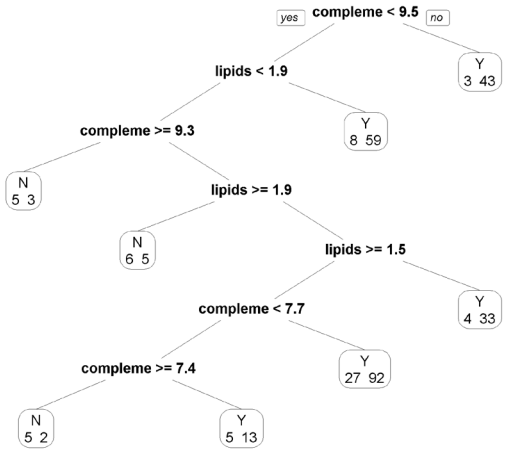
A



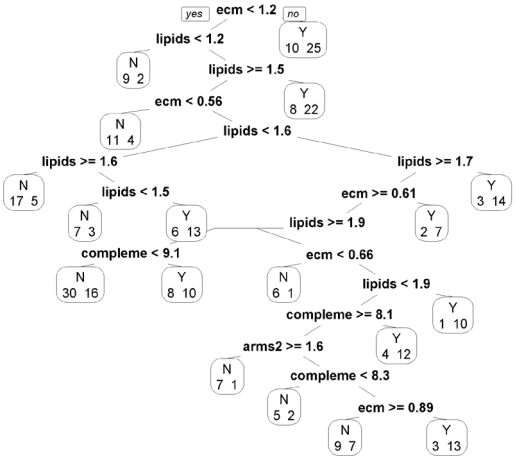
B



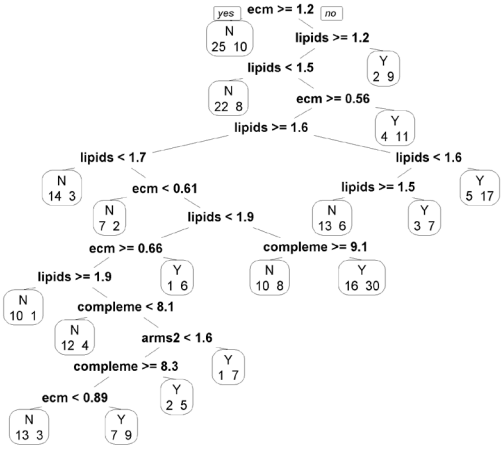
C



G



H



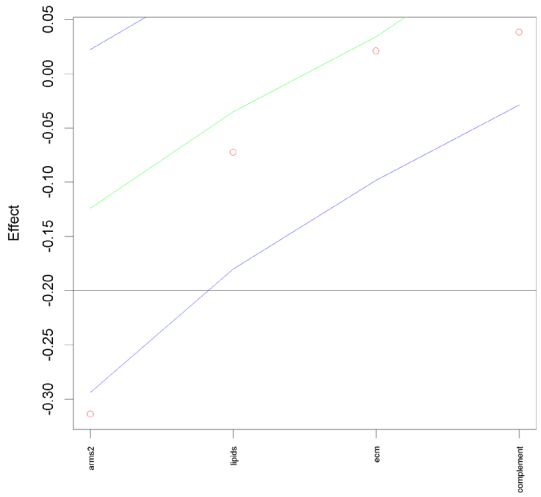
* Meaningful classification by CART was not possible for clusters D-F, owing to small numbers.

```
Summary statistics for cluster A  | Summary statistics for cluster B  |  Summary statistics for cluster C
Confusion matrix                  | Confusion matrix                  |  Confusion matrix
Predicted   N   Y                 | Predicted   N   Y                 |  Predicted   N   Y
        N  61  20                 |         N 180  52                 |          N  16  10
        Y  52 180                 |         Y  20  61                 |          Y  47 240
Accuracy = 0.769968               | Accuracy = 0.769968               |  Accuracy = 0.817891
Sensitivity = 0.900000            | Sensitivity = 0.539823            |  Sensitivity = 0.960000
Specificity = 0.539823            | Specificity = 0.900000            |  Specificity = 0.253968
----------------------------------+-----------------------------------+-----------------------------------
```

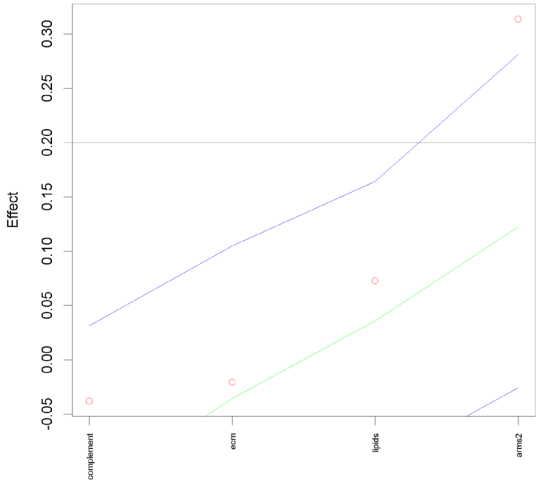
```
Summary statistics for cluster G
Confusion matrix
predicted   N   Y
        N 101  41
        Y  45 126
 Accuracy = 0.725240
 Sensitivity = 0.754491
 Specificity = 0.691781
```

Supplementary Figure 15. Cohen's effect sizes for the clusters identified by phenotypic characteristics, based on effect sizes for the genetic characteristics.

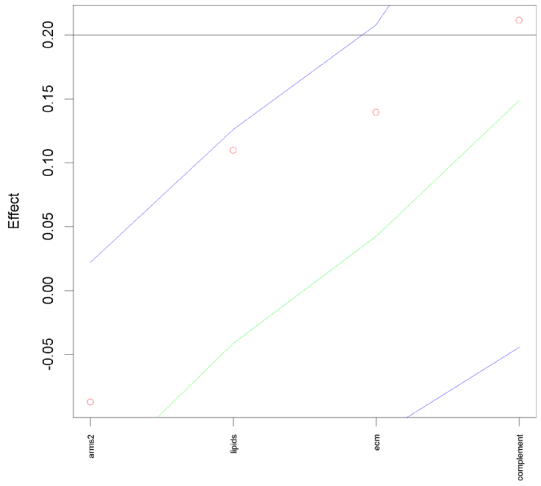
Cluster A



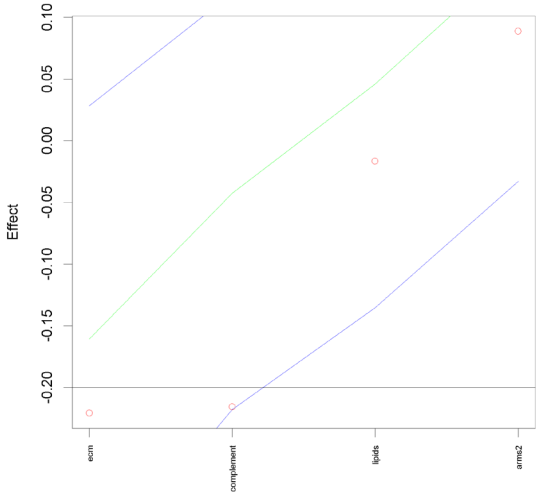
Cluster B



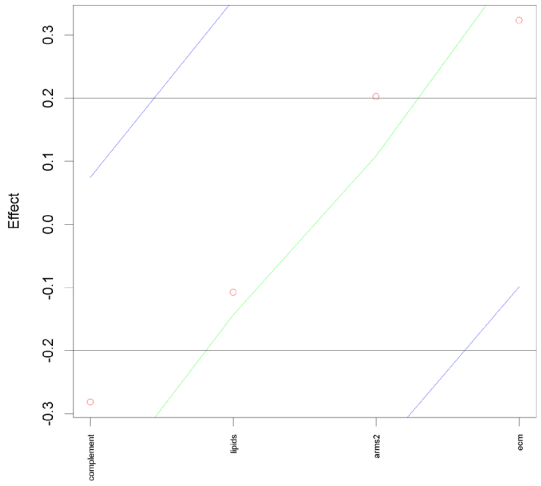
Cluster C



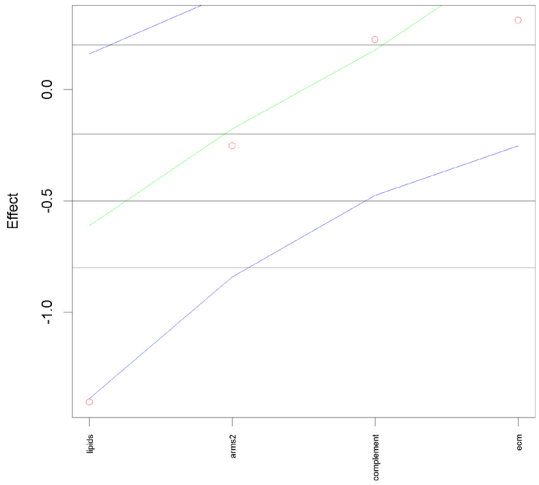
Cluster D



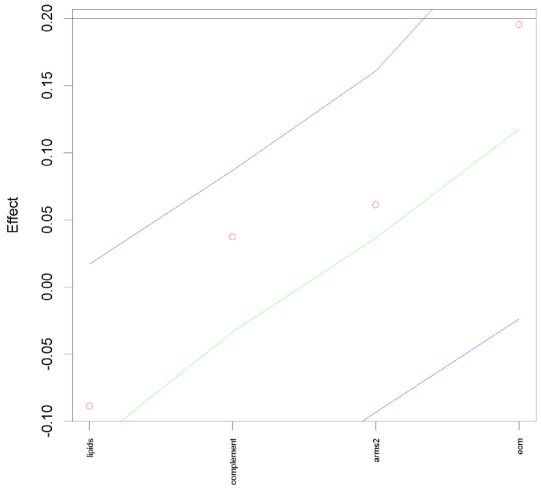
Cluster E



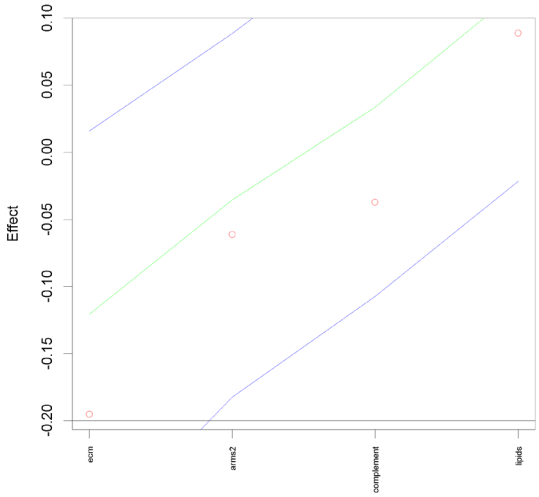
Cluster F



Cluster G



Cluster H

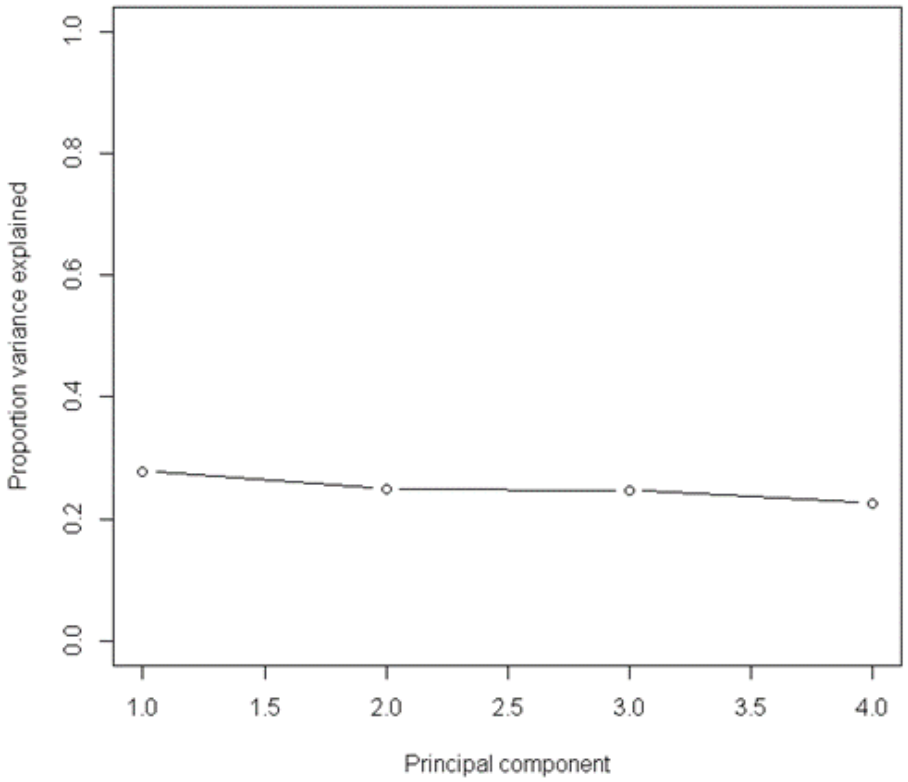


The effect sizes are plotted in ascending order on the y axis. Horizontal reference lines mark effect sizes for +/-0.2, +/-0.5, and +/-0.8. Points are colored red or black according to whether they are for Cohen's d (continuous features) or Cohen's h (binary features). The green and blue curves are expected values and 95% confidence intervals from a simulation of 1000 iterations under the null hypothesis that the cluster is unrelated to the features.

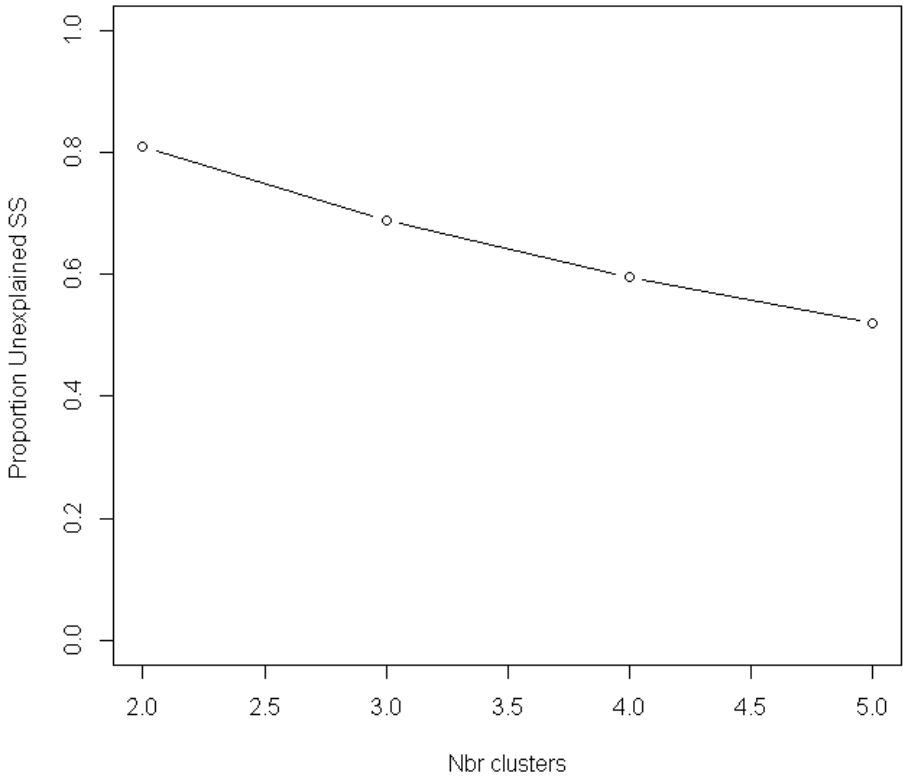
In all figures, the dots are mostly inside the blue intervals, suggesting that there is no strong evidence that the genetic variables are closely related to the phenotypic clusters.

Note: the logistic regression with LASSO either failed to converge or the predicted classification for every participant was either always inside (for clusters A, C, and G) or outside (for clusters B, D, and H). Essentially, the method failed to use the genetic data to classify the participants in a way similar to the clusters.

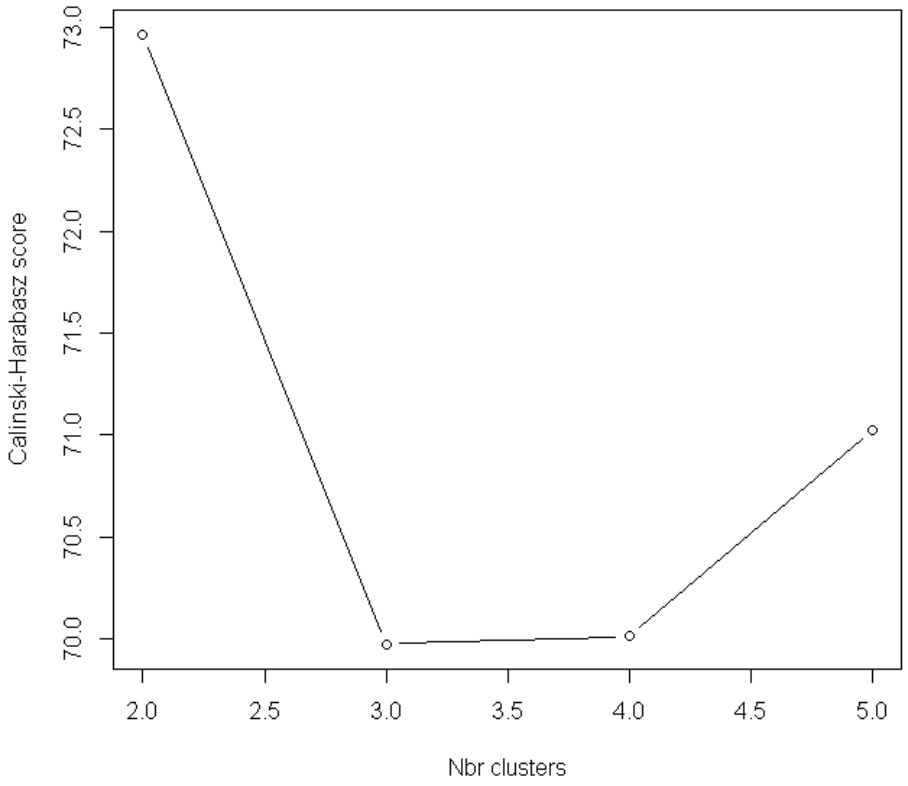
Supplementary Figure 16. Scree plot showing the proportion of variance explained by each principal component, following principal components analysis applied to the dataset of genetic characteristics.



Supplementary Figure 17. Plot showing the proportion of the unexplained sums of squares according to cluster number, following k-means cluster analysis applied to the dataset of genetic characteristics.

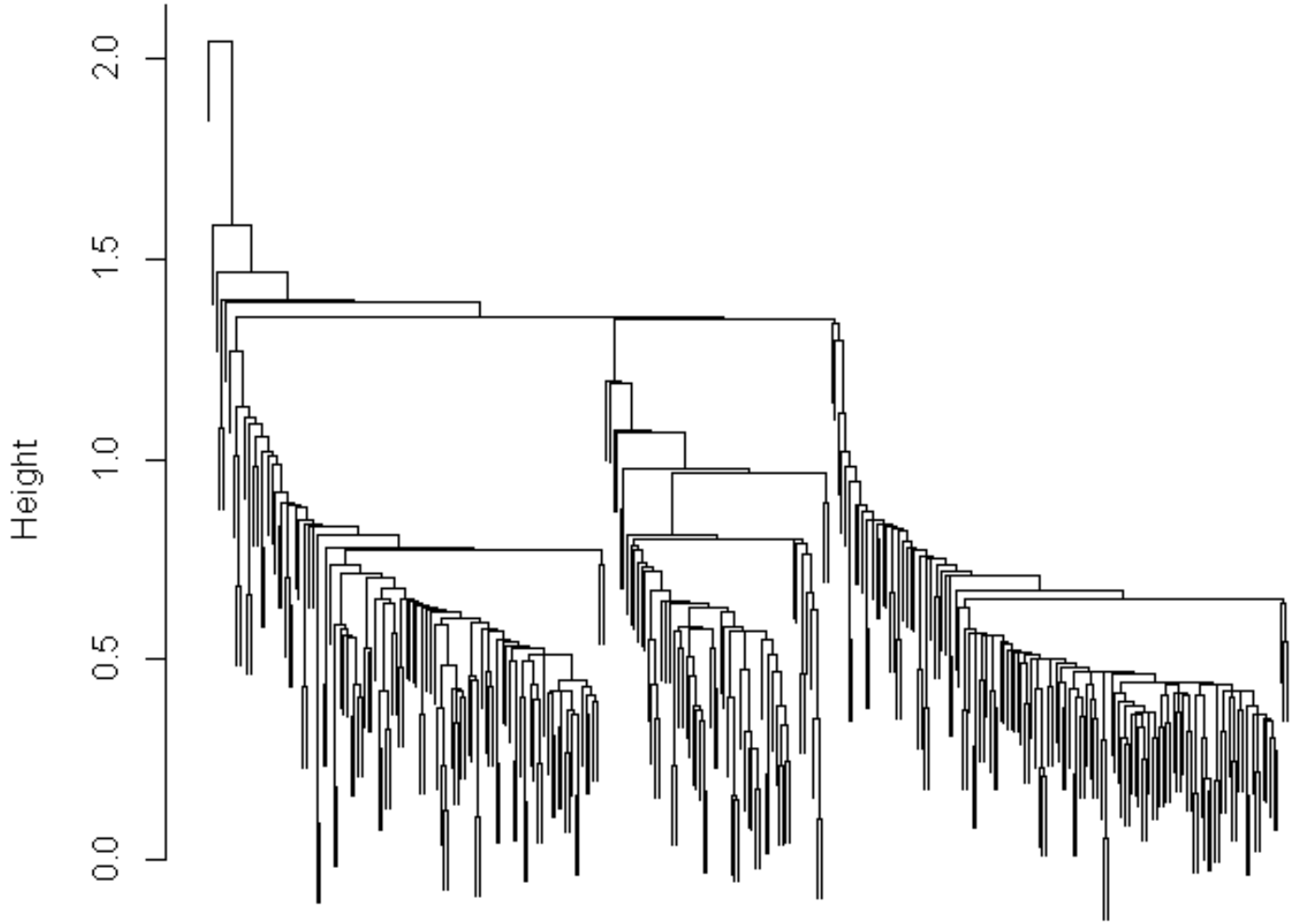


Supplementary Figure 18. Plot showing the Calinski-Harabasz scores according to cluster number, following k-means cluster analysis applied to the dataset of genetic characteristics.

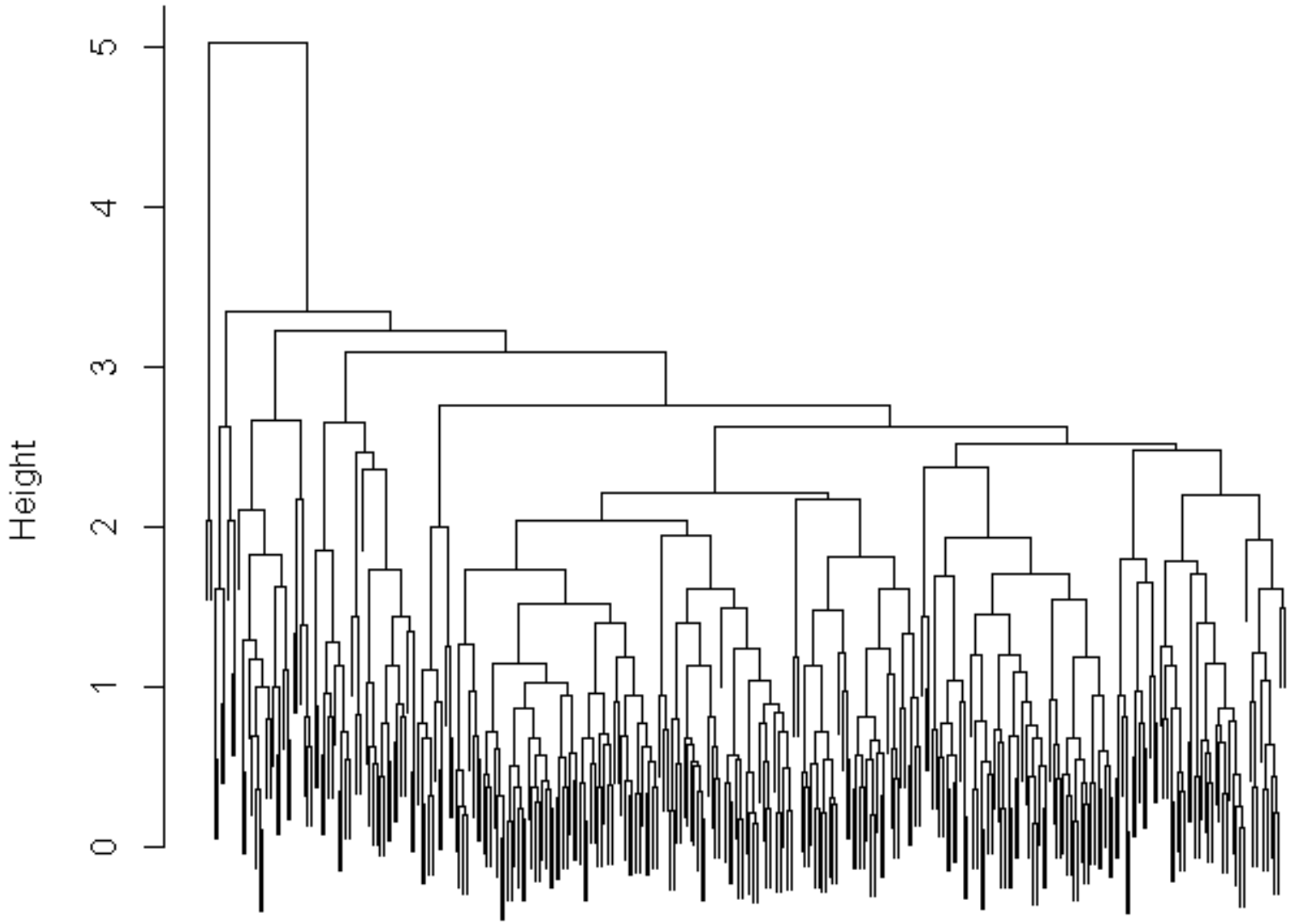


Supplementary Figure 19. Dendrograms following agglomerative hierarchical cluster analysis applied to the dataset of genetic characteristics, according to linkage type. Dissimilarity is plotted on the y-axis and each participant is shown on the x-axis. Each horizontal line represents the fusion of a pair of clusters, with the height of the segment showing the dissimilarity between the members of the pair. Clusters that fuse near the bottom of the tree are more similar, while clusters that fuse near the top are less similar.

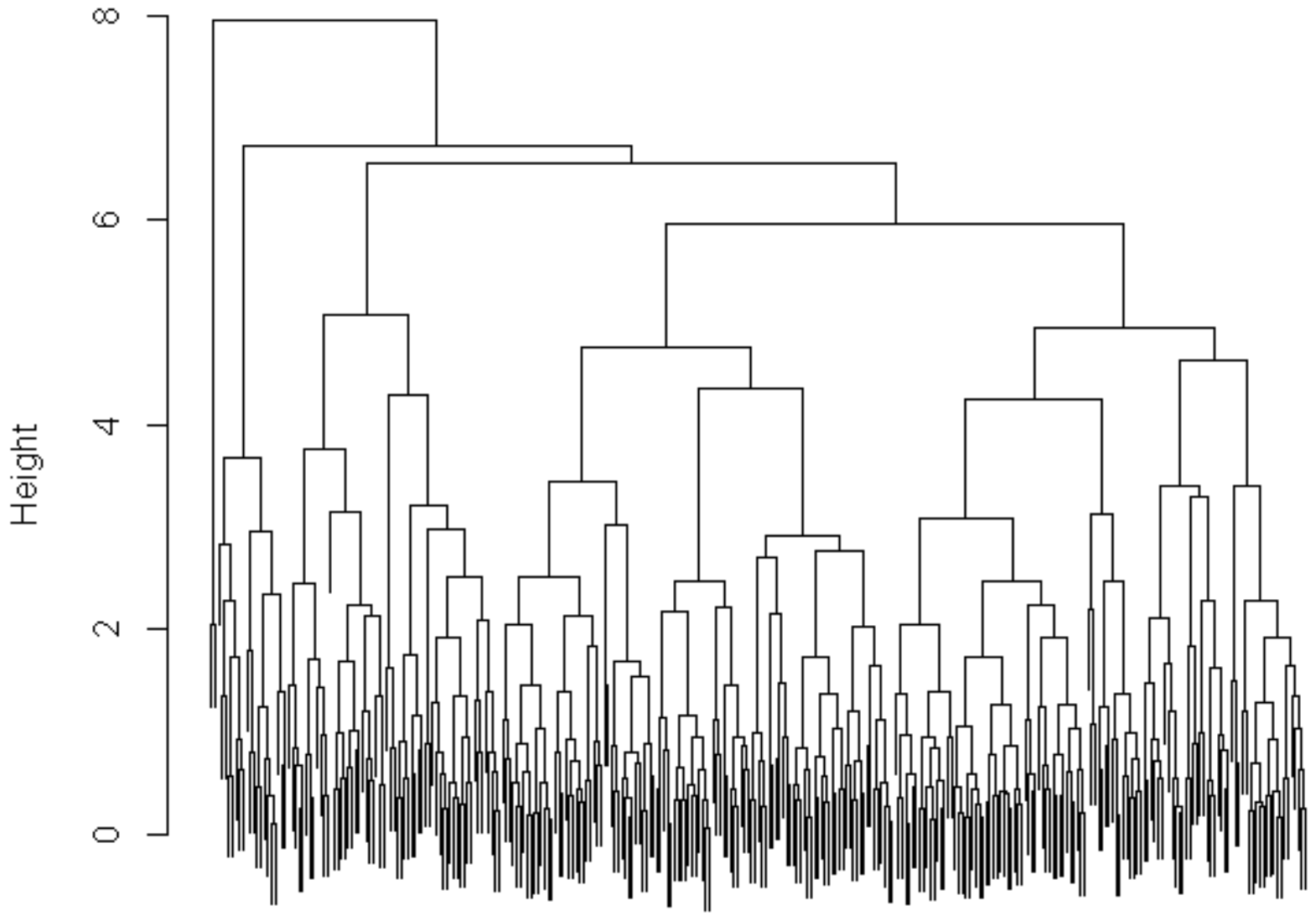
Single linkage (cophenetic correlation coefficient 0.49)



Average linkage (cophenetic correlation coefficient 0.62)

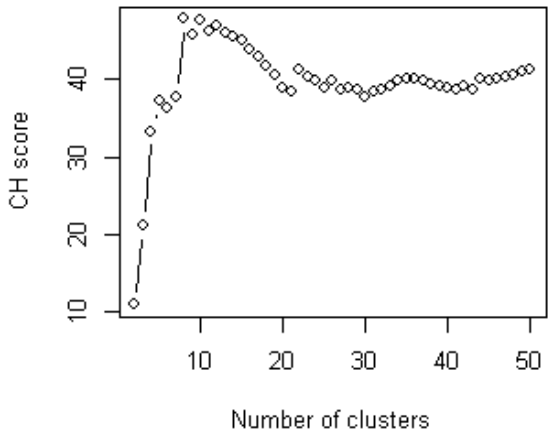


Complete linkage (cophenetic correlation coefficient 0.48)

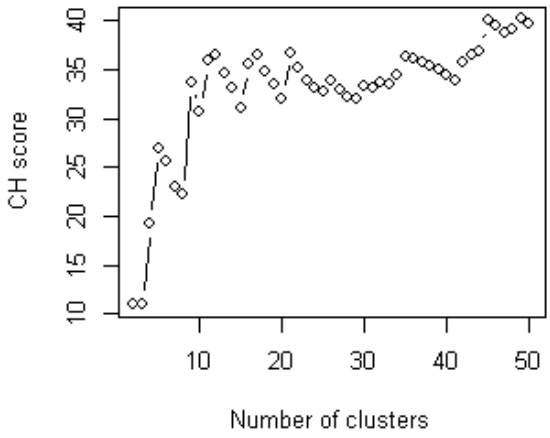


Supplementary Figure 20. Plot showing the Calinski-Harabasz scores according to cluster number, following agglomerative hierarchical cluster analysis applied to the dataset of genetic characteristics, according to linkage type.

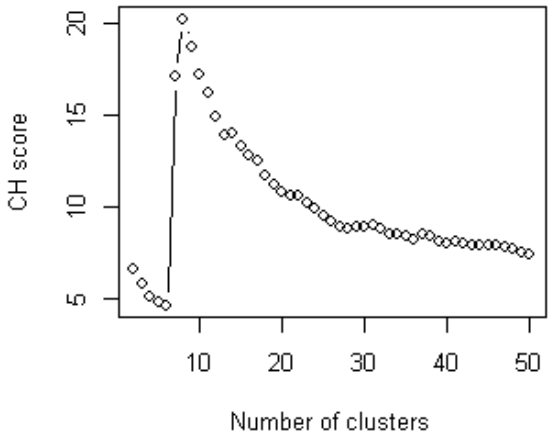
Complete linkage



Average linkage



Single linkage



Supplementary Table 2. Demographic, phenotypic, and genetic characteristics of the genetic clusters.

Variable		Cluster K (N=106)	Cluster L (N=135)	Cluster M (N=19)	Cluster N (N=11)	Cluster O (N=11)	Cluster P (N=16)	Cluster Q (N=7)	Cluster R (N=6)	Cluster S (N=2)
Age (years): median (IQR)		75.1 (70.0, 79.3)	75.1 (69.3, 79.5)	73.8 (64.4, 77.4)	72.4 (71.4, 77.4)	77.5 (73.1, 80.0)	71.4 (65.3, 76.3)	66.5 (58.0, 77.3)	74.2 (70.6, 77.3)	67.0 (65.7, 68.2)
Male: n (%)		50 (47.2)	49 (36.3)	10 (52.6)	4 (36.4)	3 (27.3)	6 (37.5)	1 (14.3)	2 (33.3)	1 (50.0)
White: n (%)		106 (100.0)	135 (100.0)	16 (84.2)	11 (100.0)	11 (100.0)	16 (100.0)	7 (100.0)	6 (100.0)	2 (100.0)
Education: n (%)	High School or Less	34 (32.1)	48 (35.6)	7 (36.8)	5 (45.5)	5 (45.5)	4 (25.0)	1 (14.3)	2 (33.3)	1 (50.0)
	At least some College	54 (50.9)	63 (46.7)	8 (42.1)	3 (27.3)	6 (54.5)	9 (56.3)	2 (28.6)	2 (33.3)	1 (50.0)
	Post-graduate	18 (17.0)	24 (17.8)	4 (21.1)	3 (27.3)	0 (0.0)	3 (18.8)	4 (57.1)	2 (33.3)	0 (0.0)
Smoking status: n (%)	Current	4 (3.8)	8 (5.9)	2 (10.5)	0 (0.0)	0 (0.0)	0 (0.0)	2 (28.6)	0 (0.0)	0 (0.0)
	Former	52 (49.1)	75 (55.6)	9 (47.4)	8 (72.7)	4 (36.4)	10 (62.5)	3 (42.9)	5 (83.3)	1 (50.0)
	Never	50 (47.2)	52 (38.5)	8 (42.1)	3 (27.3)	7 (63.6)	6 (37.5)	2 (28.6)	1 (16.7)	1 (50.0)
Central GA: n (%)	Yes	40 (37.7)	45 (33.3)	4 (21.1)	4 (36.4)	2 (18.2)	3 (18.8)	1 (14.3)	0 (0.0)	1 (50.0)
Calcified Drusen: n (%)	Yes	42 (39.6)	57 (42.2)	5 (26.3)	8 (72.7)	5 (45.5)	7 (43.8)	3 (42.9)	4 (66.7)	1 (50.0)
GA Configuration: n (%)	Small (single patch <1DA)	76 (71.7)	74 (54.8)	13 (68.4)	6 (54.5)	10 (90.9)	11 (68.8)	4 (57.1)	5 (83.3)	1 (50.0)
	Multifocal	18 (17.0)	35 (25.9)	2 (10.5)	3 (27.3)	1 (9.1)	2 (12.5)	1 (14.3)	1 (16.7)	1 (50.0)
	Horseshoe, Ring	1 (0.9)	4 (3.0)	1 (5.3)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
	Solid (center or not)	10 (9.4)	20 (14.8)	3 (15.8)	2 (18.2)	0 (0.0)	3 (18.8)	1 (14.3)	0 (0.0)	0 (0.0)
	Indeterminate	1 (0.9)	2 (1.5)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	1 (14.3)	0 (0.0)	0 (0.0)
Drusen Area Within the ETDRS Grid: n (%)	Definite, < circle C1	1 (0.9)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
	Definite, < circle C2	2 (1.9)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
	Definite, < circle I2	0 (0.0)	3 (2.2)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
	Definite, < circle O2	5 (4.7)	5 (3.7)	1 (5.3)	0 (0.0)	0 (0.0)	0 (0.0)	1 (14.3)	0 (0.0)	1 (50.0)
	Definite, < 1/2 DA	25 (23.6)	22 (16.3)	3 (15.8)	2 (18.2)	3 (27.3)	3 (18.8)	2 (28.6)	0 (0.0)	0 (0.0)
	Definite, < 1 DA	20 (18.9)	37 (27.4)	5 (26.3)	6 (54.5)	0 (0.0)	6 (37.5)	2 (28.6)	5 (83.3)	0 (0.0)

Variable		Cluster K (N=106)	Cluster L (N=135)	Cluster M (N=19)	Cluster N (N=11)	Cluster O (N=11)	Cluster P (N=16)	Cluster Q (N=7)	Cluster R (N=6)	Cluster S (N=2)
	Definite, >= 1 DA	53 (50.0)	68 (50.4)	10 (52.6)	3 (27.3)	8 (72.7)	7 (43.8)	2 (28.6)	1 (16.7)	1 (50.0)
Maximum Drusen Size: n (%)	Definite, <63 um (circle C0)	1 (0.9)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
	Definite, <125 um (circle C1)	3 (2.8)	1 (0.7)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	1 (14.3)	0 (0.0)	0 (0.0)
	Definite, <250 um (circle C2)	45 (42.5)	69 (51.1)	11 (57.9)	1 (9.1)	4 (36.4)	7 (43.8)	5 (71.4)	2 (33.3)	1 (50.0)
	Definite, >=250 um (circle C2)	57 (53.8)	65 (48.1)	8 (42.1)	10 (90.9)	7 (63.6)	9 (56.3)	1 (14.3)	4 (66.7)	1 (50.0)
GA in Fellow Eye: n (%)		26 (24.5)	31 (23.0)	7 (36.8)	0 (0.0)	2 (18.2)	2 (12.5)	1 (14.3)	0 (0.0)	2 (100.0)
RPD score: median (IQR)		0.28 (0.06, 0.69)	0.37 (0.13, 0.80)	0.19 (0.08, 0.66)	0.77 (0.04, 0.89)	0.28 (0.11, 0.85)	0.40 (0.14, 0.84)	0.22 (0.05, 0.43)	0.48 (0.12, 0.69)	0.22 (0.13, 0.31)
Square Root of GA area (mm): median (IQR)		0.8 (0.6, 1.2)	0.9 (0.6, 1.3)	0.7 (0.6, 1.3)	1.1 (0.9, 2.0)	0.8 (0.6, 1.1)	0.8 (0.6, 1.1)	1.3 (0.6, 1.8)	0.6 (0.6, 1.0)	1.1 (0.9, 1.3)
GA Enlargement from Regression of Square Root of GA area (mm/year): median (IQR)		0.19 (0.08, 0.37)	0.33 (0.08, 0.55)	0.18 (0.11, 0.30)	0.33 (0.24, 0.51)	0.18 (0.10, 0.29)	0.13 (0.04, 0.22)	0.48 (0.36, 0.64)	0.05 (0.02, 0.59)	0.62 (0.21, 1.03)
Visual acuity (ETDRS letters): median (IQR)		76.0 (66.0, 83.0)	75.0 (64.0, 81.0)	74.0 (72.0, 87.0)	67.0 (48.0, 73.0)	71.0 (60.0, 85.0)	78.0 (72.0, 82.5)	75.0 (71.0, 84.0)	79.0 (77.0, 81.0)	80.5 (73.0, 88.0)
52 SNP-based Genetic Risk Score: median (IQR)		15.1 (14.3, 15.8)	16.0 (15.0, 16.6)	15.6 (14.3, 16.4)	16.4 (16.1, 16.8)	13.2 (12.2, 14.5)	17.1 (16.4, 17.3)	14.5 (13.8, 14.9)	15.5 (14.9, 16.1)	19.0 (18.8, 19.2)
Complement GRS: median (IQR)		9.0 (8.3, 9.5)	8.5 (8.1, 9.0)	9.2 (8.7, 9.7)	8.7 (8.3, 9.3)	8.0 (6.8, 8.3)	8.7 (8.5, 9.2)	6.4 (6.3, 7.1)	8.5 (8.0, 8.8)	12.7 (12.0, 13.3)
Extracellular matrix GRS: median (IQR)		0.9 (0.8, 1.0)	0.8 (0.7, 0.9)	1.0 (0.9, 1.1)	0.8 (0.8, 0.9)	0.5 (0.5, 0.7)	1.2 (1.2, 1.3)	1.0 (0.9, 1.2)	1.5 (1.5, 1.6)	0.9 (0.8, 0.9)
Lipid metabolism GRS: median (IQR)		1.7 (1.5, 1.8)	1.8 (1.7, 1.9)	1.2 (1.1, 1.3)	1.3 (1.2, 1.4)	1.9 (1.7, 2.0)	1.8 (1.7, 1.9)	1.9 (1.6, 2.1)	1.8 (1.8, 1.9)	2.1 (1.9, 2.3)
***ARMS2* GRS: median (IQR)**		0.0 (0, 0)	1.1 (1.1, 2.1)	1.1 (0, 1.1)	2.1 (2.1, 2.1)	0.0 (0, 0)	2.1 (1.1, 2.1)	1.1 (1.1, 2.1)	0.0 (0, 1.1)	0.0 (0, 0)

Supplementary Table 3. Results: p-values for pairwise comparisons of the genetic clusters, according to phenotypic characteristics, by t test.

	I vs J		K vs L	
Phenotypic characteristic	Raw	Adjusted*	Raw	Adjusted*
Age	0.71	1.00	0.63	1.00
Sex	0.052	1.00	0.088	1.00
White / non-white	0.067	1.00	1.00	1.00
Educational level 1 (y/n)	0.92	1.00	0.57	1.00
Educational level 2 (y/n)	0.65	1.00	0.51	1.00
Educational level 3 (y/n)	0.48	1.00	0.88	1.00
Smoking level 1 (y/n)	0.42	1.00	0.18	1.00
Smoking level 2 (y/n)	0.66	1.00	0.32	1.00
Smoking level 3 (y/n)	0.42	1.00	0.45	1.00
Square root of GA area (mm)	0.26	1.00	0.38	1.00
GA central involvement (y/n)†	0.86	1.00	0.47	1.00
GA configuration level 1 (y/n)†	0.014	0.64	0.007	0.35
GA configuration level 2 (y/n)†	0.23	1.00	0.088	1.00
GA configuration level 3 (y/n)†	0.49	1.00	0.26	1.00
GA configuration level 4 (y/n)†	0.13	1.00	0.21	1.00
GA configuration level 5 (y/n)†	0.37	1.00	0.71	1.00
GA fellow eye involvement (y/n)†	0.70	1.00	0.77	1.00
Square root of GA enlargement rate (mm/year)	0.020	0.86	0.023	0.95
Total drusen area within AREDS grid (7 levels)†	0.50	1.00	0.26	1.00
Maximum drusen size within AREDS grid (4 levels)†	0.95	1.00	0.82	1.00
Calcified drusen presence (y/n)†	0.58	1.00	0.69	1.00
Reticular pseudodrusen score (0.0-1.0)†	0.79	1.00	0.23	1.00
BCVA (ETDRS letter score)†	0.18	1.00	0.23	1.00
GA central involvement (y/n)‡	0.27	1.00	0.41	1.00
GA configuration level 1 (y/n)‡	0.008	0.37	0.023	0.96
GA configuration level 2 (y/n)‡	0.75	1.00	0.64	1.00
GA configuration level 3 (y/n)‡	0.077	1.00	0.34	1.00
GA configuration level 4 (y/n)‡	0.003	0.16	0.083	1.00
GA configuration level 5 (y/n)‡	0.62	1.00	0.88	1.00
GA fellow eye involvement (y/n)‡	0.52	1.00	0.39	1.00
Total drusen area within AREDS grid (7 levels)‡	0.48	1.00	0.10	1.00
Maximum drusen size within AREDS grid (4 levels)‡	0.18	1.00	0.20	1.00
Calcified drusen presence (y/n)‡	0.93	1.00	0.70	1.00
Reticular pseudodrusen score (0.0-1.0)‡	0.72	1.00	0.38	1.00
BCVA rate (change in ETDRS letter score/year)‡	0.14	1.00	0.96	1.00

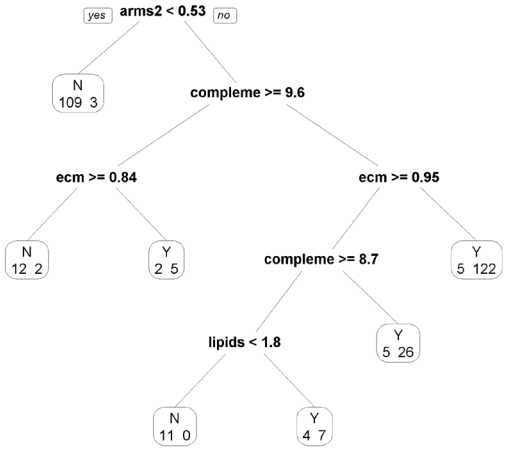
* Adjusted for multiple testing: adjusted for the 35 phenotypic characteristics by MULTTEST bootstrap and adjusted for the 2 cluster groupings by multiplying by 2.

† considered cross-sectionally

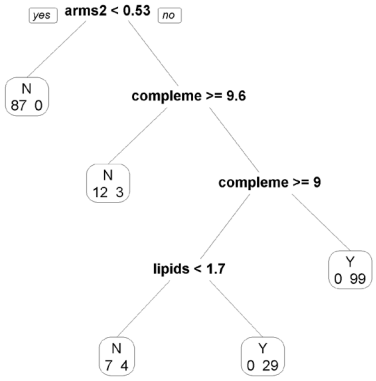
‡ considered longitudinally (as defined in Table 1)

Supplementary Figure 21. CART classification trees and related confusion matrices and performance metrics for the clusters identified by genetic characteristics, based on CART classification by the same genetic characteristics.

I vs J



K vs L



Cluster I vs J	Cluster K vs L
Confusion matrix	Confusion matrix
predicted N Y	predicted N Y
N 132 5	N 106 7
Y 16 160	Y 0 128
Accuracy = 0.932907	Accuracy = 0.970954
Sensitivity = 0.969697	Sensitivity = 0.948148
Specificity = 0.891892	Specificity = 1.000000

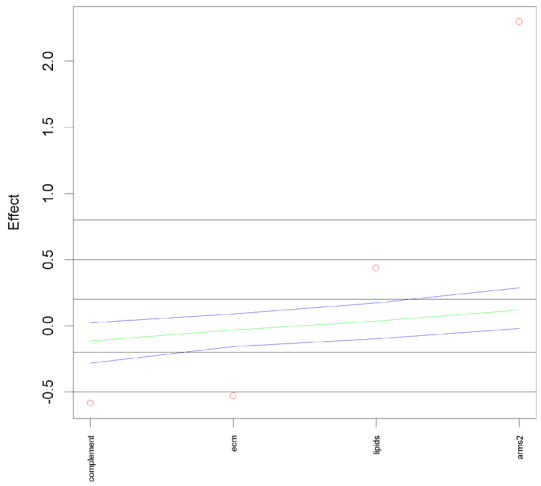
Supplementary Figure 22. Results of logistic regression with LASSO, with related confusion matrices and performance metrics, for the clusters identified by genetic characteristics, based on CART classification by the same genetic characteristics.

Log odds ratio estimates from logistic regression with LASSO

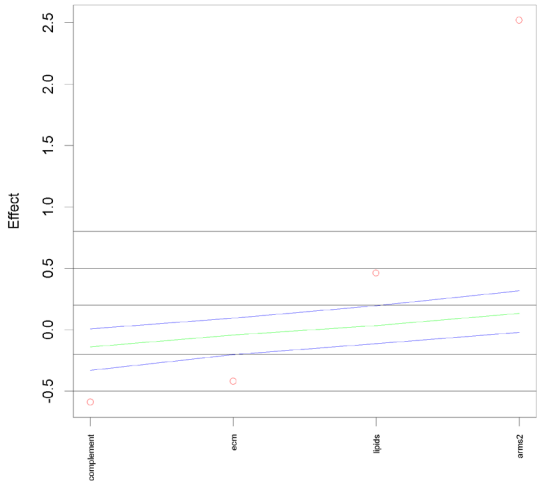
```
Coefficients for cluster I vs J    |  cluster K vs L
(Intercept)  31.783035             |  (Intercept)  3.137322
complement   -7.010674            |  complement  -1.156121
lipids       19.859494             |  lipids       2.188386
ecm         -30.966055             |  ecm          .
arms2        24.786952             |  arms2        4.188201
                                   |
Confusion matrix                   |  Confusion matrix
pred   N   Y                       |  pred   N   Y
   N 148   0                       |     N  97   0
   Y   0 165                       |     Y   9 135
Accuracy = 1.000000                |  Accuracy = 0.962656
Sensitivity = 1.000000             |  Sensitivity = 1.000000
Specificity = 1.000000             |  Specificity = 0.915094
```

Supplementary Figure 23. Cohen's effect sizes for the clusters identified by genetic characteristics, based on CART classification by the same genetic characteristics.

Cluster I vs J



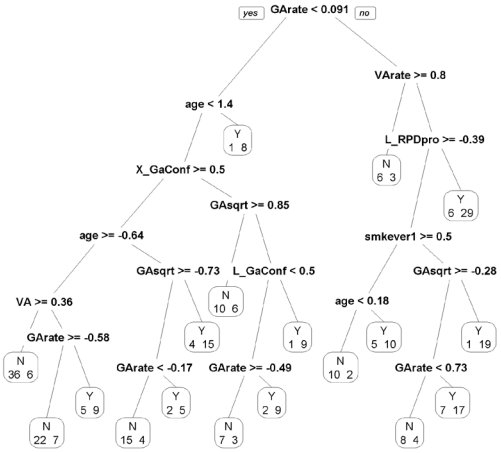
Cluster K vs L



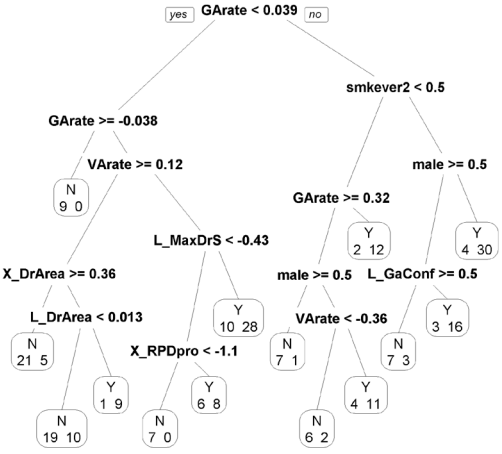
The effect sizes are plotted in ascending order on the y axis. Horizontal reference lines mark effect sizes for +/-0.2, +/-0.5, and +/-0.8. Points are colored red or black according to whether they are for Cohen's d (continuous features) or Cohen's h (binary features). The green and blue curves are expected values and 95% confidence intervals from a simulation of 1000 iterations under the null hypothesis that the cluster is unrelated to the features.

Supplementary Figure 24. CART classification trees and related confusion matrices and performance metrics for the clusters identified by genetic characteristics, based on CART classification by the phenotypic characteristics.

I vs J



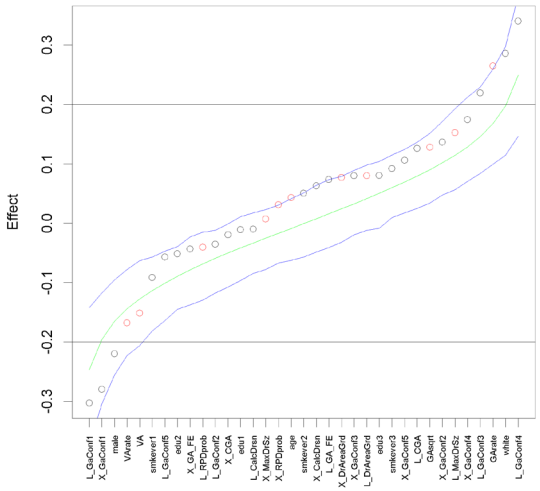
K vs L



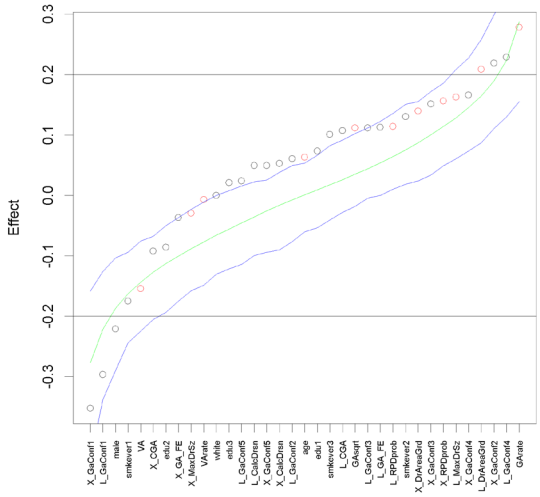
Cluster I vs J	Cluster K vs L
Confusion matrix	Confusion matrix
predicted N Y	predicted N Y
N 114 35	N 76 21
Y 34 130	Y 30 114
Accuracy = 0.779553	Accuracy = 0.788382
Sensitivity = 0.787879	Sensitivity = 0.844444
Specificity = 0.770270	Specificity = 0.716981

Supplementary Figure 25. Cohen's effect sizes for the clusters identified by genetic characteristics, based on CART classification by the phenotypic characteristics.

Cluster I vs J



Cluster K vs L



The effect sizes are plotted in ascending order on the y axis. Horizontal reference lines mark effect sizes for +/-0.2, +/-0.5, and +/-0.8. Points are colored red or black according to whether they are for Cohen's d (continuous features) or Cohen's h (binary features). The green and blue curves are expected values and 95% confidence intervals from a simulation of 1000 iterations under the null hypothesis that the cluster is unrelated to the features.

In all figures, the dots are mostly inside the blue intervals, suggesting that there is no strong evidence that the phenotypic variables are closely related to the genetic clusters.

Note: the logistic regression with LASSO failed. Specifically, all coefficients except the intercept term were set to 0, such that the predicted classification for every participant was either always inside or always outside the cluster. Essentially, the method failed to use the phenotypic data to classify the participants in a way similar to the clusters.