CLUSTER ANALYSIS IDENTIFIES DISTINCT PATHOGENETIC PATTERNS IN C3 GLOMERULOPATHIES/IMMUNE COMPLEX–MEDIATED MEMBRANOPROLIFERATIVE GN

Paraskevas Iatropoulos, Erica Daina, Manuela Curreri, Rossella Piras, Elisabetta Valoti, Caterina Mele, Elena Bresin, Sara Gamba, Marta Alberti, Matteo Breno, Annalisa Perna, Serena Bettoni, Ettore Sabadini, Luisa Murer, Marina Vivarelli, Marina Noris and Giuseppe Remuzzi

SUPPLEMENTAL MATERIAL

SUPPLEMENTAL METHODS

Clinical, histologic, biochemical and genetic data

Clinical data were recorded using a standardized Case Report Form. All the biopsy reports were independently reviewed by two pathologists and discordances were resolved by face-to-face discussion. Immunostaining was performed by immunofluorescence in all cases.

Serum C3 and C4 levels were assessed by kinetic nephelometry.[1] Plasma SC5b-9 levels were measured using the MicroVue SC5b-9 Plus EIA commercial kit (SC5b-9 Plus, Quidel). C3NeF activity was determined by purifying IgG from plasma and assessing their ability to stabilize cell-bound C3bBb convertase as previously described.[2]

Targeted Next-generation sequencing for the genetic screening of all exons and flanking regions of *CFH*, *CD46*, *CFI*, *CFB*, *C3* and *THBD* genes was performed by highly multiplex PCR using the Ion AmpliSeq™ Library Kit 2.0 followed by template preparation and sequencing on an Ion PGM Sequencer as previously described.[3] Variants fulfilling the following criteria were considered as likely pathogenetic variants: 1. variants previously identified in patients with complement disorders and with functional assays supporting variant pathogenicity; 2. variants with allele frequency ≤0.001 in any subpopulation of the ExAC database and CADD score ≥10.

Likely pathogenetic variants segregated with the disease in all familial cases when the affected relatives were tested (n = 5). As for penetrance, we have analyzed 38 unaffected relatives of sporadic and familial cases and found 12 healthy carriers of the likely pathogenetic variants. The penetrance was estimated to be 57% (16 affected out of 28 LPV carriers).

Variable reduction and Clustering

We used 34 different variables (Supplemental Table 1): 7 for clinical features at onset, 17 for pathologic findings (7 for light microscopy, 6 for IF and 4 for EM), 3 regarding the serum or plasma complement profile, 1 regarding genetic or acquired complement abnormalities and 6 regarding common SNPs of complement genes.

To remove noisy variables,[4] we performed a principal component analysis including all the variables and we retained significant components (i.e. eigenvalue ≥1), since components with eigenvalue <1 contribute little to explain the relationships between original variables.[4] Principal component analysis was performed with the PCAmixdata package (on R platform), that enables the analysis for a mixture of quantitative and qualitative variables. Successively, an unsupervised hierarchical clustering was performed based on significant components using the Ward's method with Euclidean distances with the stats package on R. For the cluster analysis, imputation of the missing values was performed with the *k-Nearest Neighbour* method and iterative robust regression using the VIM package. Patients and variables with >15% of missing data were removed.

We have tested various measures to estimate the optimal number of cluster (NbClust package for R platform). Of 23 solutions 7 suggested 2 clusters, 3 suggested 3 clusters, 6 suggested 4 clusters, 1 suggested 5 clusters, and 1, 1, 3 and 1 suggested 7, 8, 9 and 10 clusters, respectively. As shown in Supplemental Figure 2 the results from Hartigan's rule justified up to four clusters. Ultimately, we prioritized the four cluster solution to achieve a good balanced data distribution as well as relevance to the goal of this study to find out distinct disease entities characterized by specific pathophysiological mechanisms (the two cluster solution had 73% of patients in one cluster).

To check the stability of the clustering, we used a home-made script that randomly half-split the study population and repeated the same clustering method. This procedure was repeated 10,000 times and the case memberships were cross-checked with that of the original clustering each time. Cohen's

Kappa agreement rates were all in the moderate range (overall 0.47 and 0.42, 0.40, 0.56, 0.51 for clusters 1, 2, 3 and 4, respectively; Supplemental Table 4).[5]

To cross-validate the results of the unsupervised hierarchical clustering by a different approach, we performed the *kmeans* clustering technique with the kmeans++ method of initialization using the ClusterR package on the R platform.

Identification of Criteria for patient grouping

For ease of use, quantitative variables were dichotomized by adopting cutoffs. The selection of features that could be used to identify clusters was performed in three phases as previously described.[6] First, the odds ratio was calculated for each feature with the presence or absence of a cluster as dependent variable and the features with statistically significant odds ratios were included in the next phase. Second, logistic regression in a backward stepwise procedure was carried out for each cluster separately as dependent variable and the features that were significantly associated with one of the clusters were selected for the next phase. Third, multinomial logistic regression was performed with clusters as the dependent variable and the features from phase two as the independent variables and those non reaching significance were discarded. This model contained all the finally selected features together. Multinomial logistic regression was performed with the *mlogit* package on the R platform.

In the first step of the algorithm, we assigned different points to each feature and we summed the points for each patient to obtain a score. ROC curve was drawn for this score to determine the optimal cutoff to distinguish clusters.

Statistical analyses

We used the ANOVA test for continuous variables and the χ^2 test (or Fisher Exact test, where appropriate) for categorical variables. Survival analyses considered as cumulative fractions of patients free of events were estimated using Kaplan-Meier survival curves and Cox proportional-Hazards analysis. Only variables with a *p*-value <0.05 in the univariate analysis were considered in the multivariate analysis. *P*-values <0.05 were considered statistically significant.

Supplemental Table 1. Patients with likely pathogenic variants present in the cohort.

Patient ID	Likely pathogenic variant	Zyg.	Histol. group	Cluster	Group	ExAC Global Freq.	ExAC Max Subpop. Freq.	Pathogenic in functional studies	CADD	C3NeF	Serum C3 (mg/dl)	Plasma SC5-9 (ng/ml)	New/ known
1549	CFH:R2I (chr1:196621252G>T)	Het	DDD	3	3	0	0	No	11	Yes	54	286	(3)
2032	CFH:R78G (chr1:196642281A>G)	Hom	C3GN	1	1	0	0	No	16	No	15	1530	(3,7-9)
2158	CFH:R78G (chr1:196642281A>G)	Hom	C3GN	1	1	0	0	No	16	No	14	4571	(3,7-9)
1287	CFH:P88T (chr1:196643004C>A)	Hom	IC	2	2	0	0	No	29	No	48	2074	(3)
1284	CFH:P88T (chr1:196643004C>A)	Hom	IC	2	2	0	0	No	29	No	5	3596	(3)
2082	CFH:R127C (chr1:196645147C>T)	Het	C3GN	1	1	0	0	No	33	NA	72	2789	(3)
2192	CFH:G133R (chr1:196645165G>A)	Het	C3GN	1	1	1.7E-05	3.0E-05	No	31	NA	55	355	New
1073	CFH:C494R (chr1:196683008T>C)	Het	IC	4	2	0.0E+00	0.0E+00	No	24	No	70	1332	New
419	CFH:Y1008X (chr1:196711070delTA)	Hom	C3GN	3	1	0	0	No	32	No	18	NA	(3)
1101	CFH:R1210C (chr1:196716375C>T)	Het	DDD	1	3	1.7E-04	2.8E-04	Yes	12	No	154	368	(1,3, 10,11)
2018	CD46:K66N (chr1:207930459A>T)	Het	IC	2	2	5.3E-04	9.0E-04	No	13	No	63	466	(3)
1360	CFI:c.1-4C>T (chr4:110723131G>A)	Het	C3GN	1	1	8.4E-06	1.5E-05	No	17	Yes	17	399	(3)
1409	CFI:G57D (chr4:110687868C>T)	Het	DDD	3	3	8.2E-06	6.1E-05	No	25	Yes	47	321	(3)
SN249	CFI G119R (chr4:110685820C>T)	Het	C3GN	1	1	5.3E-04	9.5E-04	Yes	22	NA	79	310	(1,3,11)
1264	CFB:G161R (chr6:31914966G>A)	Het	IC	2	1	6.9E-05	3.0E-04	No	27	Yes	36	1600	(3)
1147	CFB:H451R (chr6:31917278A>G)	Het	IC	1	2	0	0	No	25	Yes	40	1874	(3)
1726	CFB:R679W (chr6:31919196C>T)	Hom	IC	1	2	0	0	No	31	No	20	291	(3)
216	CFI:R317W (chr4:110670750G>A)	Het	C3GN	1	1	9.9E-05	2.3E-04	No	16	NA	18	NA	(3,12)
1964	C3:R505C (chr19:6710823G>A)	Het	IC	2	1	8.3E-06	1.5E-05	No	25	Yes	18	495	(3)
1828	C3:V619M (chr19:6707931C>T)	Het	C3GN	1	1	2.9E-04	0.001	No	22	No	57	561	(3)
1890	C3:G637R (chr19:6707877C>G)	Het	IC	1	2	2.2E-04	3.8E-04	No	24	Yes	12	1845	(3)
1983	C3:I761del (chr19:6702553delGAT)	Het	C3GN	1	1	0	0	No	14	No	33	845	New
1984	C3:I761del (chr19:6702553delGAT)	Het	C3GN	1	1	0	0	No	14	No	38	888	New
216	C3:R1042Q (chr19:6694471C>T)	Het	C3GN	1	1	0	0	No	32	NA	18	NA	(3,7)
521	C3:K1051M (chr19:6694444T>A)	Het	IC	4	1	1.6E-05	3.0E-05	Yes	23	NA	41	348	(3,13)
2047	C3:S1063N (chr19:6693465C>T)	Het	IC	1	2	1.1E-04	8.1E-04	Yes	10	No	70	235	(3)
1736	C3:R1303H (chr19:6685060C>T)	Het	C3GN	1	1	8.3E-06	1.5E-05	No	28	No	25	413	(3)

1132	C3:R1320Q (chr19:6685009C>T)	Het	C3GN	1	1	0	0	No	28	No	15	1087	(3)
2011	C3:D1362N (chr19:6684607C>T)	Het	IC	3	4	4.9E-05	1.2E-04	No	11	NA	112	337	(3)
2508	C3:D1456N (chr19:6680259C>T)	Het	C3GN	1	1	0	0	No	25	No	10	1844	New
2009	C3:C1518R (chr19:6679214A>G)	Het	C3GN	2	2	0	0	No	26	Yes	6	2520	(3)
1157	C3:D1625H (chr19:6678010G>C)	Het	C3GN	1	1	0	0	No	12	Yes	2	3750	(3)
1741	THBD:D486Y (chr20:23028686G>T)	Het	C3GN	1	1	0	0	Yes	12	Yes	40	1298	(3,14)
1725	THBD:P495S (chr20:23028659G>A)	Het	DDD	3	3	5.8E-04	0.001	Yes	6	Yes	18	137	(3,14)

Zyg.: zygosity; Histol.Group: Histologic group according to the current classification; Algor. cluster: algoritm-based cluster; ExAC Global Freq.: variant frequency in all subjects of the ExAC database (v0.3); ExAC Max Subpop. Freq.: variant frequency in non-Finnish European subjects of the ExAC database (v0.3). Path. in functional studies: Functional studies supporting pathogenicity are available. CADD: CADD Phred score (v1.3). In brackets the reference of the variants already described in MPGN, C3G or aHUS patients. IC: immune-complex mediated membranoproliferative glomerulonephritis. Serum C3: reference 90-180 mg/dl; serum C4: reference 10-40 mg/dl; plasma SC5b-9: reference ≤400 ng/ml.

Supplemental Table 2. Patients examined and subdivided according to cluster analysis.

Patient ID	Histol. group	Cluster	Algor. cluster	Age of onset (y)	Likely Pathog. Variants	C3NeF	Serum C3 (mg/dl)	Plasma SC5-9 (ng/ml)	Glomer. C1q	Intr. Dense deps
216	C3GN	1	1	3.5	Yes	NA	18	NA	Neg	No
419	C3GN	3	1	0.7	Yes	No	18	NA	Neg	No
521	IC	4	1	13.0	Yes	NA	41	348	Neg	No
633	DDD	1	3	11.3	No	Yes	4	1364	Trace	Yes
923	DDD	3	3	8.1	No	Yes	10	337	Trace	Yes
949	IC	1	1	11.2	No	Yes	20	303	Neg	No
1001	IC	4	4	71.4	No	No	63	438	2+	No
1020	C3GN	4	4	13.1	No	Neg	76	NA	Neg	No
1026*	IC	2	2	7.0	No	Yes	5	1080	1+/2+	No
1055	IC	4	2	56.9	No	Yes	40	NA	3+	No
1073	IC	4	2	28.0	Yes	No	70	1332	2+	No
1101	DDD	1	3	48.8	Yes	No	154	368	Neg	Yes
1115	DDD	3	3	14.0	No	No	49	237	Neg	Yes
1124	DDD	3	3	9.9	No	Yes	20	1906	Trace	Yes
1132	C3GN	1	1	18.2	Yes	No	15	1087	Neg	No
1135	DDD	3	3	44.7	No	No	28	571	Neg	Yes
1147	IC	1	2	18.6	Yes	Yes	40	1874	1+	No
1149	DDD	3	3	9.7	No	Yes	9	674	Neg	Yes
1157	C3GN	1	1	9.6	Yes	Yes	2	3750	Trace	No
1169	C3GN	1	1	13.3	No	Yes	10	1766	Neg	No
1192	IC	1	1	28.4	No	Pos	76	840	Neg	No
1194	IC	2	2	44.9	No	Yes	15	2922	2+	No
1218	DDD	3	3	12.3	No	Yes	55	188	Neg	Yes
1228	IC	4	2	11.0	No	No	10	682	2+	No
1240	C3GN	4	4	18.6	No	Yes	112	462	Neg	No
1261	C3GN	4	4	10.6	No	No	107	195	Neg	No
1264	IC	2	1	5.3	Yes	Yes	36	1600	Neg	No
1267	IC	4	4	26.9	No	No	77	281	1+/2+	No
1279	IC	2	2	11.4	No	No	41	642	2+	No
1284	IC	2	2	0.4	Yes	No	5	3596	1+	No
1286	DDD	1	3	6.0	No	Yes	12	1890	Neg	Yes
1287	IC	2	2	0.3	Yes	No	48	2074	1+/2+	No
1304	IC	2	2	30.6	No	Yes	45	249	2+	No
1330	IC	3	3	11.0	No	NA	50	NA	2+	Yes
1332	DDD	3	3	18.2	No	NA	5	173	Neg	Yes
1357	C3GN	1	1	13.4	No	NA	3	NA	Neg	No
1359	IC	2	4	4.3	No	No	125	354	1+/2+	No
1360	C3GN	1	1	1.8	Yes	Yes	17	399	Trace	No
1370	IC	4	4	32.7	No	No	63	226	Neg	No
1402	C3GN	1	1	11.3	No	Yes	76	1987	Neg	No
1406	C3GN	1	1	12.0	No	Yes	45	983	Trace	No
1409	DDD	3	3	9.5	Yes	Yes	47	321	Neg	Yes
1411	IC	4	4	17.5	No	No	74	291	Neg	No
1416	DDD	3	3	13.5	No	Yes	36	552	Neg	Yes
1419	IC	2	2	10.9	No	Yes	5	2462	3+	No
1428	IC	4	4	30.4	No	No	107	307	Neg	No

1453	IC	1	2	9.3	No	No	7	1093	1+	No
1468	DDD	3	3	10.0	No	NA	12	NA	Trace	Yes
1490	IC	3	1	9.6	No	Yes	8	467	Neg	No
1491	C3GN	3	1	8.0	No	Yes	48	206	Neg	No
1492	C3GN	4	4	5.8	No	Yes	111	89	Neg	No
1498	C3GN	1	1	4.2	No	Yes	40	149	Neg	No
1499	C3GN	1	1	10.2	No	Yes	14	764	Neg	No
1540	C3GN	3	1	5.3	No	No	16	359	Neg	No
1545	C3GN	4	4	22.3	No	No	75	196	Neg	No
1549	DDD	3	3	24.7	Yes	Yes	54	286	Neg	Yes
1553	C3GN	4	4	17.5	No	Neg	112	NA	Neg	No
1558	IC	4	4	41.3	No	No	107	755	1+	No
1567	IC	4	1	16.8	No	Yes	68	2120	Neg	No
1594	C3GN	4	1	54.0	No	No	27	286	Neg	No
1605	C3GN	4	4	17.0	No	No	179	651	Neg	No
1606	IC	1	4	11.4	No	No	90	242	Neg	No
1625	IC	1	2	10.4	No	Yes	16	1433	1+	No
1633	C3GN	1	4	19.3	No	No	57	486	1+	No
1676	C3GN	1	4	2.6	No	No	77	169	Neg	No
1678	DDD	3	3	26.0	No	No	45	269	Neg	Yes
1680	C3GN	2	2	38.4	No	No	44	229	1+	No
1684	IC	4	4	55.0	No	No	63	346	2+	No
1685	C3GN	4	4	59.7	No	No	132	186	Neg	No
1706	IC	1	2	9.9	No	No	17	3523	2+	No
1707	C3GN	4	4	38.3	No	No	99	217	Neg	No
1710	IC	1	1	14.9	No	Yes	11	4488	Neg	No
1724	IC	2	4	0.7	No	No	80	248	Neg	No
1725	DDD	3	3	12.3	Yes	Yes	18	137	Neg	Yes
1726	IC	1	2	24.8	Yes	No	20	291	1+	No
1727	C3GN	4	4	9.2	No	No	53	273	Neg	No
1736	C3GN	1	1	17.9	Yes	No	25	413	Neg	No
1741	C3GN	1	1	8.0	Yes	Yes	40	1298	Neg	No
1742*	DDD	3	3	11.7	No	Yes	5	380	1+	Yes
1750	IC	2	4	45.3	No	No	76	364	3+	No
1763	IC	1	2	11.4	No	Yes	17	1391	1+	No
1769	IC	2	2	15.0	No	No	9	3353	1+	No
1773	DDD	3	3	11.8	No	Yes	9	267	Neg	Yes
1804	IC	2	2	17.0	No	Yes	17	2474	3+	No
1805	IC	4	4	0.5	No	No	156	171	Neg	No
1825	C3GN	4	4	4.0	No	No	154	207	Neg	No
1826	IC	4	4	13.0	No	No	107	268	1+	No
1828	C3GN	1	1	27.5	Yes	No	57	561	Neg	No
1830	C3GN	4	4	22.0	No	No	111	204	Neg	No
1834	C3GN	1	1	13.4	No	Yes	6	5517	Neg	No
1837*	DDD	3	3	10.6	No	Yes	9	545	1+	Yes
1849	IC	1	2	16.3	No	Yes	13	2086	2+	No
1852	IC	4	4	32.0	No	No	102	469	1+	No
1853	IC	4	4	15.0	No	No	103	164	Neg	No
1857	C3GN	1	1	22.0	No	No	21	932	Neg	No

1872	IC	2	4	24.9	No	No	147	265	Neg	No
1875	C3GN	1	1	10.2	No	Yes	44	772	Neg	No
1876	IC	2	2	41.9	No	No	18	601	1+	No
1882	C3GN	1	1	8.2	No	Yes	4	2709	Neg	No
1883	IC	1	1	3.0	No	Yes	5	1175	Neg	No
1886	IC	2	2	22.0	No	Yes	8	NA	1+	No
1890	IC	1	2	10.0	Yes	Yes	12	1845	2+	No
1891	IC	2	2	14.0	No	Yes	14	167	3+	No
1895	C3GN	1	1	6.4	No	Yes	42	1015	Neg	No
1927	IC	2	2	41.0	No	No	17	534	3+	No
1933	C3GN	1	1	17.0	No	Yes	50	656	Neg	No
1949	IC	2	2	17.0	No	Yes	4	2390	2+	No
1961	C3GN	1	1	3.8	No	No	46	156	Neg	No
1964	IC	2	1	12.0	Yes	Yes	18	495	Neg	No
1967	IC	3	3	22.0	No	NA	84	257	1+	Yes
1968	IC	3	1	12.0	No	Yes	51	278	Trace	No
1970	IC	4	4	10.0	No	No	54	220	Neg	No
1979	IC	2	4	15.0	No	No	NA	329	1+	No
1980	IC	2	4	6.0	No	Yes	111	268	2+	No
1983	C3GN	1	1	22.0	Yes	No	33	845	Neg	No
1984	C3GN	1	1	19.5	Yes	No	38	888	Neg	No
1985	C3GN	4	1	9.6	No	No	17	417	Neg	No
1986	IC	2	2	11.0	No	Yes	4	3365	1+	No
2008	C3GN	3	1	13.0	No	Yes	21	462	Neg	No
2009	C3GN	2	2	15.2	Yes	Yes	6	2520	1+	No
2011	IC	3	4	42.8	Yes	Neg	112	337	Neg	No
2012	C3GN	1	2	11.0	No	Yes	6	2184	1+	No
2014	C3GN	1	2	20.0	No	Yes	5	4490	1+	No
2018	IC	2	2	21.7	Yes	No	63	466	3+	No
2020	DDD	3	3	7.5	No	Yes	23	1310	Neg	Yes
2024	IC	4	4	4.5	No	No	94	195	3+	No
2032	C3GN	1	1	24.0	Yes	No	15	1530	Neg	No
2033	IC	2	2	12.0	No	Yes	3	3526	2+	No
2038	IC	4	4	37.0	No	No	114	254	1+	No
2040	IC	4	4	36.0	No	No	61	654	Neg	No
2042	C3GN	1	1	20.7	No	No	49	NA	Neg	No
2047*	IC	1	2	10.0	Yes	No	70	235	1+	No
2050	C3GN	1	1	4.0	No	NA	18	NA	Neg	No
2070	C3GN	1	1	17.0	No	Pos	NA	NA	Neg	No
2081	IC	2	2	8.5	No	Yes	16	1643	2+	No
2082	C3GN	1	1	0.8	Yes	NA	72	2789	Neg	No
2109	DDD	3	3	14.0	No	No	36	223	Neg	Yes
2116	IC	2	2	9.2	Yes	Yes	10	3520	1+	No
2117	C3GN	4	4	6.0	No	No	130	165	1+	No
2118	IC	4	4	25.0	No	No	99	363	Neg	No
2121	IC	2	2	12.0	No	Yes	45	593	1+/2+	No
2125	IC	4	4	45.0	No	No	80	664	3+	No
2158	C3GN	1	1	41.7	Yes	No	14	4571	Neg	No
2162	DDD	3	3	4.9	No	Yes	16	195	Neg	Yes

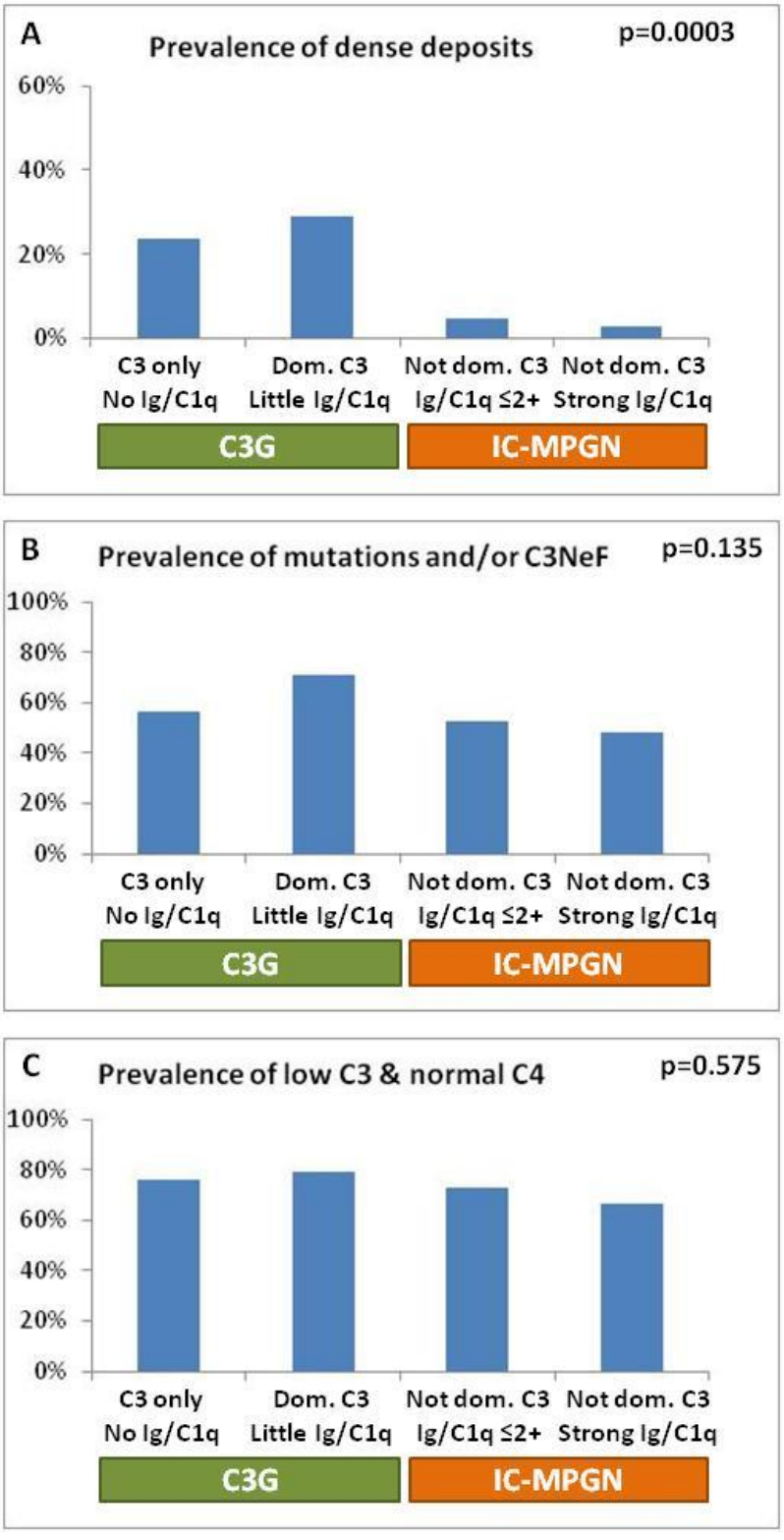
2163*	IC	3	3	6.5	No	Yes	18	277	Neg	Yes
2164	C3GN	1	2	0.6	No	Yes	8	3424	1+	No
2167	C3GN	4	4	23.0	No	No	89	319	Neg	No
2173	C3GN	1	1	15.5	No	No	48	1134	Neg	No
2177	IC	1	4	8.6	No	No	63	248	1+	No
2191	C3GN	1	1	47.0	No	No	27	959	Neg	No
2192	C3GN	1	1	9.0	Yes	NA	55	355	Neg	No
2198	DDD	1	3	8.0	No	Yes	110	479	Neg	Yes
2227	C3GN	4	4	62.0	No	No	124	209	1+	No
2246	C3GN	4	4	72.0	No	No	81	494	Neg	No
2255	IC	4	4	50.7	No	No	63	237	1+	No
2256	C3GN	1	1	46.0	No	No	21	4207	Trace	No
2257	C3GN	1	1	15.0	No	Yes	24	2270	Neg	No
2258	C3GN	1	1	16.0	No	Pos	77	874	Neg	No
2279	IC	1	1	52.0	No	No	31	166	Neg	No
2288	C3GN	4	1	27.0	No	No	36	1550	Neg	No
2297	DDD	3	3	6.0	No	Yes	3	167	Neg	Yes
2298	IC	2	2	19.0	No	No	18	2438	2+	No
2299	IC	3	1	6.0	No	No	18	223	Neg	No
2300	C3GN	1	1	10.0	No	No	18	1601	Neg	No
2301	DDD	3	3	28.0	No	Yes	76	236	Neg	Yes
2333	C3GN	4	4	60.6	No	No	109	29	Neg	No
2343	IC	2	2	16.3	No	No	15	3521	2+	No
2421	IC	4	4	34.9	No	No	83	281	2+	No
2466	IC	4	4	11.9	No	No	84	206	1+/2+	No
2467	IC	4	4	56.7	No	No	73	485	2+	No
2487	IC	3	2	34.5	No	Yes	15	5880	2+	No
2508	C3GN	1	1	11.5	Yes	No	10	1844	Neg	No
SN249	C3GN	1	1	2.0	Yes	NA	79	310	Neg	No

Histol.Group: histologic group according to the current classification; Algor. cluster: algoritm-based cluster; Likely Pathog. Variants: likely pathogenetic variants; Glomer. C1q: glomerular C1q staining; Intr. Dense deps: intramembranous highly electron-dense deposits; Serum C3: reference 90-180 mg/dl; serum C4: reference 10-40 mg/dl; plasma SC5b-9: reference ≤400 ng/ml. *Positive for anti-Factor H autoantibodies.

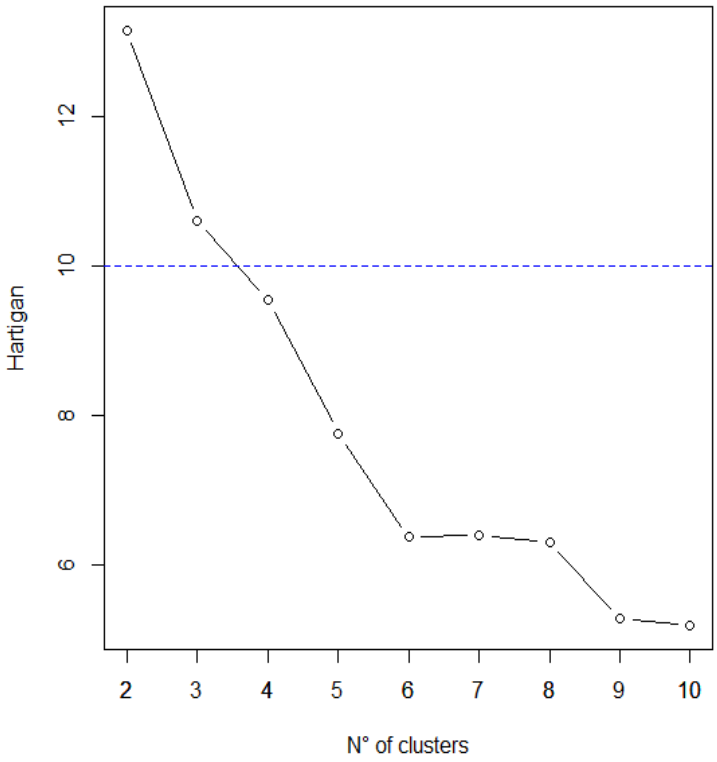
Supplemental Table 3. Variables included in the principal component analysis.

Variable	**Codification**
Gender (male/female)	0: male; 1: female
Age at onset	years
Hematuria at onset	0: absence of hematuria; 1: microhematuria; 2: gross hematuria
Proteinuria at onset	0: absence of proteinuria; 1: proteinuria; 2: nephrotic syndrome
Renal impairment at onset	0: absence of renal impairment; 1: renal impairment; 2: ESRD
Trigger event at onset	0: No; 1: Yes
Familiarity for nephropathy	0: No; 1: Yes
% Sclerotic glomeruli on LM	%
% of crescents on LM	%
Degree of mesangial proliferation on LM	0 to 3+ scale
Degree of endocapillary proliferation on LM	0 to 3+ scale
Degree of interstitial inflammation on LM	0 to 3+ scale
Degree of interstitial fibrosis on LM	0 to 3+ scale
Degree of arteriolar sclerosis on LM	0 to 3+ scale
C3 on IF	0 to 3+ scale
IgA on IF	0 to 3+ scale
IgG on IF	0 to 3+ scale
IgM on IF	0 to 3+ scale
C1q on IF	0 to 3+ scale
Fibrinogen on IF	0 to 3+ scale
Mesangial deposits on EM	0: No; 1: Yes
Subepithelial deposits on EM	0: No; 1: Yes
Subendothelial deposits on EM	0: No; 1: Yes
Intramembranous highly electron-dense ribbon-like deposits on EM	0: No; 1: Yes
Serum C3	0: very low (<50mg/dl); 1: low (≥50 & <90 mg/dl); 2: normal (≥90 mg/dl)
Serum C4	0: low (<10mg/dl); 1: lower normal range (≥10 & <20 mg/dl); 2: medium & higher normal range (≥20 mg/dl)
Plasma SC5b-9 (ng/ml)	0: normal (≤303ng/ml); 1: high (>303 & ≤800 ng/ml); 2: very high (>800 mg/dl)
Number of alternative pathway abnormalities(including Likely Pathogenic Variants and C3NeF)	0, 1, 2, …
CFH V62I genotype	N° of I alleles
CFH Y402H genotype	N° of H alleles
CD46 c.-366A>G genotype	N° of G alleles
CFB Q/W32R genotypes	N° of R alleles
C3 R102G genotype	N° of G alleles
THBD A473V genotype	N° of V alleles

Supplemental Figure 1. Prevalence of intramembranous highly electron-dense deposits (**panel A**), complement gene likely pathogenetic variants (LPVs) and/or C3NeF (**panel B**) and low serum C3 with normal C4 (**panel C**). C3G and IC-MPGN patients are divided in four groups according to IF features: bright C3 only with no immunoglobulin (Ig) or C1q staining, dominant C3 staining (C3 ≥2 orders of magnitude than Ig and C1q) with traces or 1+ staining in a least one of Ig or C1q, not dominant C3 staining (C3 <2 orders of magnitude than Ig and C1q) with maximum 2+ staining in Ig or C1q, and not dominant C3 staining with 3+ staining in at least one of Ig or C1q.



Supplemental Figure 2. Hartigan number calculated according to the Hartigan's rule for each potential number of clusters illustrating 4 as the best choice.



Supplemental Table 4. Cohen's Kappa agreement rates between the clusters and the validation approaches, and between the clusters and the algoritms based clusters.

	Overall	Cluster 1	Cluster 2	Cluster 3	Cluster 4
Half-splitting (10,000 iterations)	0.467	0.419	0.401	0.562	0.508
Kmeans approach	0.568	0.434	0.610	0.830	0.467
Algorithm-based clusters	0.646	0.563	0.580	0.742	0.724

Cohen's Kappa agreement rates: < 0 indicating no agreement, 0–0.20 slight, 0.21–0.40 fair, 0.41–0.60 moderate, 0.61–0.80 substantial, and 0.81–1 almost perfect agreement.

Supplemental Table 5. Univariate and backward multivariate binomial logistic regression analysis to identify the criteria for patients' classification adopting a significance p-value threshold of 0.05.

Cluster	Feature	Univariate analysis		Multivariate analysis	
		OR	*p*	OR	*p*
1	N° of Alternative Pathway abnormalities[c]	2.1	0.002		
1	Serum C3	2.6	3.9E-04		
1	Plasma SC5b-9	3.0	1.1E-06	3.6	1.5E-04
1	Sclerotic Glomeruli > 5%	0.44	0.024	0.24	0.015
1	Degree of interstitial inflammation ≥1	0.29	0.016		
1	Glomerular IgG deposits ≥1+	0.35	0.003	0.24	0.022
1	Glomerular C1q deposits ≥1+	0.34	0.003	0.13	0.002
1	Mesangial deposits	4.9	5.2E-05	7.6	0.001
1	Subepithelial deposits	2.8	0.002	6.8	7.5E-04
1	Subendothelial deposits	4.2	4.8E-04	15.6	7.5E-05
1	Intramembranous highly electron dense deposits	0.25	0.014		
1	Familiarity for nephropathy	2.7	0.027	4.6	0.030
1	Renal impairment at onset	0.35	0.027	0.25	0.039
1	THBD A473V	0.28	0.007	0.13	0.008
2	Serum C4	2.3	0.002		
2	Plasma SC5b-9	1.63	0.048		
2	Degree of mesangial proliferation ≥1	0.45	0.050	0.26	0.028
2	Degree of endocapillary proliferation ≥1	3.3	0.004		
2	Glomerular IgA deposits ≥1+	5.9	1.4E-04		
2	Glomerular IgG deposits ≥1+	16.0	9.5E-07	12.0	6.9E-04
2	Glomerular IgM deposits ≥1+	3.3	0.006		
2	Glomerular C1q deposits ≥1+	19.7	1.5E-07	9.1	0.001
2	Glomerular Fibrinogen deposits ≥1+	2.5	0.047	6.6	0.014
2	Subendothelial deposits	9.0	0.003		
2	Proteinuria at onset	4.7	8.3E-05	6.5	5.6E-04
2	CFH V62I	2.4	0.006	3.1	0.020
2	CFH H402Y	2.4	0.005		
3	N° of Alternative Pathway abnormalities[c]	2.1	0.011		
3	Serum C3	2.7	0.008	7.7	0.006
3	Plasma SC5b-9	0.48	0.003		
3	Sclerotic Glomeruli > 5%	0.11	0.003		
3	Crescents > 40%	37.4	8.9E-04	462	3.5E-05
3	Glomerular IgG deposits ≥1+	0.39	0.033		
3	Glomerular C1q deposits ≥1+	0.24	0.005		
3	Subepithelial deposits	0.26	0.006		
3	Subendothelial deposits	0.03	1.9E-09	0.02	7.4E-04
3	Intramembranous highly electron dense deposits	90.7	2.0E-12	127	6.9E-06
3	Hematuria at onset	2.1	0.017		

4	N° of Alternative Pathway abnormalities	0.06	8.4E-09	0.06	1.4E-04
4	Serum C3	0.12	2.0E-11	0.12	7.3E-06
4	Serum C4	0.48	0.012		
4	Plasma SC5b-9	0.36	1.3E-05		
4	Sclerotic Glomeruli > 5%	6.6	5.1E-07		
4	Degree of endocapillary proliferation ≥1	0.24	5.2E-04	0.15	0.008
4	Degree of interstitial inflammation ≥1	2.8	0.012		
4	Degree of interstitial fibrosis ≥1	3.7	0.006		
4	Degree of arteriolar sclerosis ≥1	8.5	5.5E-04		
4	Age of onset ≥18 years	3.7	2.4E-04		
4	Hematuria at onset	0.34	1.0E-04	0.24	0.004
4	Renal impairment at onset	5.0	5.1E-05	10.9	0.002
4	THBD A473V	2.5	0.008		

[c]Abnormalities include the alleles with Likely Pathogenetic Variants and C3NeF; the prevalence of patients with 1 and 2 abnormalities is reported. [d]Serum C3 subdivided as normal (≥90mg/dl), low (≥50 and <90 mg/dl) and very low (<50mg/dl); the prevalence of patients with low and very low C3 is reported. [e]Plasma SCb-9 subdivided as normal (≤303ng/ml), high (>303 and ≤800 ng/ml) and very high >800 ng/ml); the prevalence of patients with high and very high SC5b-9 is reported. [f]Hematuria subdivided as no hematuria, microhematuria and gross hematuria; prevalence of microhematuria and gross hematuria is reported. [g]Proteinuria subdivided as no proteinuria, proteinuria and nephrotic syndrome; prevalence of proteinuria and nephrotic syndrome is reported. [h]Number of the 62I alleles.

Supplemental Table 6. Multivariate multinomial regression showing the features available at onset that independently predict the clusters adopting a significance p-value threshold of 0.05.

Feature	Prevalence	Group vs. reference[a]	β	RR[b] (e^{β})	p
N° of Alternative Pathway abnormalities[c]	45% : 12%	4	-3.5	0.03	0.038
Serum C3[d]	23% : 61%	3	9.0	8.0E+03	0.025
		4	-5.7	0	0.012
Plasma SC5b-9	29% : 39%	2	-4.3	0.01	0.014
		3	-4.9	0.01	0.012
		4	-4.8	0.01	0.008
Sclerotic Glomeruli > 5%	32%	2	7.8	2405	0.001
		4	7.9	2673	0.004
Degree of mesangial proliferation ≥1	86%	2	-3.6	0.03	0.025
Degree of endocapillary proliferation ≥1	57%	2	5.0	154	0.012
Glomerular IgG deposits ≥1+	41%	2	9.9	19331	0.002
		4	5.1	168	0.040
Glomerular C1q deposits ≥1+	38%	2	9.3	11422	0.002
Glomerular Fibrinogen deposits ≥1+	16%	2	8.1	3152	0.009
Mesangial deposits	63%	2	-12.0	6.0E-06	0.001
		3	-7.9	3.7E-04	0.045
		4	-10.5	2.8E-05	0.002
Subepithelial deposits	40%	2	-4.1	0.02	0.020
		3	-13.3	1.7E-06	0.005
		4	-8.3	2.4E-04	0.002
Subendothelial deposits	68%	3	-18.8	7.0E-09	0.005
Intramembranous highly electron dense deposits	16%	3	9.5	12886	0.027
Hematuria at onset	50% : 35%	4	-7.4	6.3E-04	0.003
Proteinuria at onset	56% : 35%	2	7.3	1506	0.003
Familiarity for nephropathy	14%	3	-13.7	1.1E-06	0.028
		4	-6.9	9.6E-04	0.012
CFH V62I	32% : 4%	2	6.6	700	0.001
		3	10.2	27103	0.010

[a]The group with the greatest number of patients (cluster 1) was taken as reference group. [b]Relative risk. [c]Abnormalities include the alleles with Likely Pathogenetic Variants and C3NeF; the prevalence of patients with 1 and 2 abnormalities is reported. [d]Serum C3 subdivided as normal (≥90mg/dl), low (≥50 and <90 mg/dl) and very low (<50mg/dl); the prevalence of patients with low and very low C3 is reported. [e]Plasma SCb-9 subdivided as normal (≤303ng/ml), high (>303 and ≤800 ng/ml) and very high >800 ng/ml); the prevalence of patients with high and very high SC5b-9 is reported. [f]Hematuria subdivided as no hematuria, microhematuria and gross hematuria; prevalence of microhematuria and gross hematuria is reported. [g]Proteinuria subdivided as no proteinuria, proteinuria and nephrotic syndrome; prevalence of proteinuria and nephrotic syndrome is reported. [h]Number of the 62I alleles.

Supplemental Table 7. Univariate and backward multivariate logistic regression analysis to identify the criteria for patients' classification adopting the 0.001 significance threshold.

Cluster	Feature	Univariate analysis		Multivariate analysis	
		OR	*p*	OR	*p*
1	Serum C3	2.6	3.9E-04		
1	Plasma SC5b-9	3.0	1.1E-06	2.9	7.5E-06
1	Mesangial deposits	4.9	5.2E-05	4.4	3.7E-04
1	Subendothelial deposits	4.2	4.8E-04		
2	Glomerular IgA deposits ≥1+	5.9	1.4E-04		
2	Glomerular IgG deposits ≥1+	16.0	9.5E-07		
2	Glomerular C1q deposits ≥1+	19.7	1.5E-07	19.7	1.5E-07
2	Proteinuria at onset	4.7	8.3E-05		
3	Crescents > 40%	37.4	8.9E-04	169	2.1E-05
3	Subendothelial deposits	0.03	1.9E-09		
3	Intramembranous highly electron dense deposits	90.7	2.0E-12	186	2.1E-12
4	N° of Alternative Pathway abnormalities	0.06	8.4E-09	0.07	3.1E-05
4	Serum C3	0.12	2.0E-11	0.13	2.0E-07
4	Plasma SC5b-9	0.36	1.3E-05		
4	Sclerotic Glomeruli > 5%	6.6	5.1E-07		
4	Degree of endocapillary proliferation ≥1	0.24	5.2E-04		
4	Degree of arteriolar sclerosis ≥1	8.5	5.5E-04		
4	Age of onset ≥18 years	3.7	2.4E-04		
4	Hematuria at onset	0.34	1.0E-04	0.20	4.1E-04
4	Renal impairment at onset	5.0	5.1E-05		

Supplemental Table 8. Univariate and multivariate analysis of the variables associated with end-stage renal disease.

Feature	Univariate analysis		Multivariate analysis	
	HR	***p***	**HR**	***p***
N° of Alternative Pathway abnormalities	0.35	0.034		
Sclerotic Glomeruli (%)	70.2	<0.0001	43.6	0.011
Crescents (%)	15.1	0.006	24.1	0.017
Degree of interstitial inflammation (0 to 3+)	1.9	0.053		
Degree of interstitial fibrosis (0 to 3+)	1.9	0.017		
Degree of arteriolar sclerosis (0 to 3+)	2.0	0.010		
Age of onset (years)	1.0	0.025		
Nephrotic syndrome at onset	3.9	0.011	5.4	0.007
Renal impairment at onset	4.8	0.002		
Original Cluster 4 vs. others	6.6	0.001	6.2	0.013
Algorithm-based Cluster 4 vs. others	3.7	0.013		
C3GN vs. DDD & IC-MPGN	1.24	0.693		
DDD vs. C3GN & IC-MPGN	0.40	0.389		
IC-MPGN vs. C3GN & DDD	1.02	0.967		

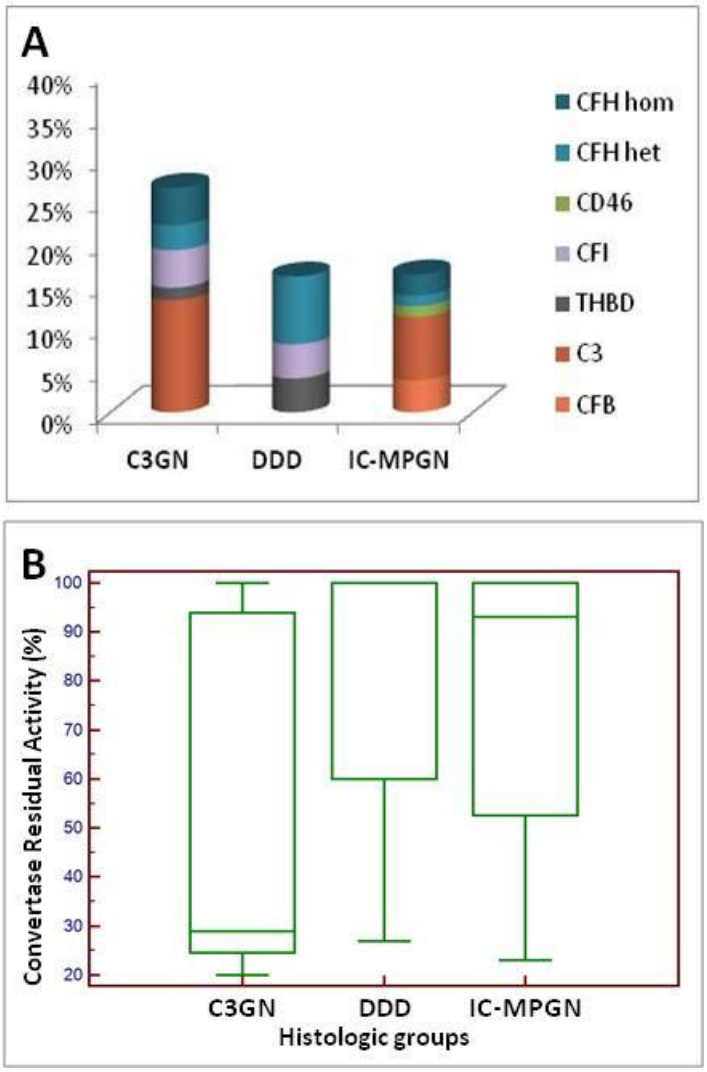
HR: Hazard ratio calculated by Multivariate Cox proportional-Hazards analysis. Nephrotic syndrome was defined as: 24-hour proteinuria exceeding 3.5g in adults or 40mg/h/m2 in children together with albuminemia ≤3 g/dL. In the multivariate analysis original clusters and algorithm-based clusters were not used at the same time, but alternatively; in the multivariate analysis, together with sclerotic glomeruli, crescents and nephrotic syndrome at onset, the algorithm-based cluster 4 *vs.* others reached a *p*=0.098 significance.

Supplemental Table 9. Treatment received by the patients subdivided according to the clusters.

	1	*2*	*3*	*4*	*p*
Original clusters					
Steroids	73%	81%	63%	80%	0.264
Intensified immunosuppression	34%	56%	43%	29%	0.077
ACEi/ARB	76%	69%	80%	73%	0.761
Algorithm-based clusters					
Steroids	77%	80%	59%	75%	0.245
Intensified immunosuppression	42%	60%	32%	21%	**0.002**
ACEi/ARB	72%	73%	84%	75%	0.682

Intensified immunosuppression:cyclosporine A, mycophenolate, tacrolimus and/or cyclophosphamide;ACEi:angiotensin-converting enzyme inhibitors;ARB: angiotensin receptor blockers.

Supplemental Figure 3. Distribution according to histologic classification of the likely pathogenetic variants (panel A) and the C3NeF residual activity evaluated by hemolytic assay in C3NeF-positive patients (panel B).



Supplemental References

1. Noris, M, Caprioli, J, Bresin, E, Mossali, C, Pianetti, G, Gamba, S, Daina, E, Fenili, C, Castelletti, F, Sorosina, A, Piras, R, Donadelli, R, Maranta, R, van der Meer, I, Conway, EM, Zipfel, PF, Goodship, TH, Remuzzi, G: Relative role of genetic complement abnormalities in sporadic and familial aHUS and their impact on clinical phenotype. *Clin J Am Soc Nephrol,* 5**:** 1844-1859, 2010.
2. Fremeaux-Bacchi, V, Weiss, L, Brun, P, Kazatchkine, MD: Selective disappearance of C3NeF IgG autoantibody in the plasma of a patient with membranoproliferative glomerulonephritis following renal transplantation. *Nephrol Dial Transplant,* 9**:** 811-814, 1994.
3. Iatropoulos, P, Noris, M, Mele, C, Piras, R, Valoti, E, Bresin, E, Curreri, M, Mondo, E, Zito, A, Gamba, S, Bettoni, S, Murer, L, Fremeaux-Bacchi, V, Vivarelli, M, Emma, F, Daina, E, Remuzzi, G: Complement gene variants determine the risk of immunoglobulin-associated MPGN and C3 glomerulopathy and predict long-term renal outcome. *Mol Immunol,* 71**:** 131-142, 2016.
4. Burgel, PR, Paillasseur, JL, Caillaud, D, Tillie-Leblond, I, Chanez, P, Escamilla, R, Court-Fortune, I, Perez, T, Carre, P, Roche, N: Clinical COPD phenotypes: a novel approach using principal component and cluster analyses. *Eur Respir J,* 36**:** 531-539, 2010.
5. Fereshtehnejad, SM, Zeighami, Y, Dagher, A, Postuma, RB: Clinical criteria for subtyping Parkinson's disease: biomarkers and longitudinal progression. *Brain,* 140**:** 1959-1976, 2017.
6. Kruizenga, HM, Seidell, JC, de Vet, HC, Wierdsma, NJ, van Bokhorst-de van der Schueren, MA: Development and validation of a hospital screening tool for malnutrition: the short nutritional assessment questionnaire (SNAQ). *Clin Nutr,* 24**:** 75-82, 2005.
7. Caprioli, J, Castelletti, F, Bucchioni, S, Bettinaglio, P, Bresin, E, Pianetti, G, Gamba, S, Brioschi, S, Daina, E, Remuzzi, G, Noris, M: Complement factor H mutations and gene polymorphisms in haemolytic uraemic syndrome: the C-257T, the A2089G and the G2881T polymorphisms are strongly associated with the disease. *Hum Mol Genet,* 12**:** 3385-3395, 2003.
8. Noris, M, Galbusera, M, Gastoldi, S, Macor, P, Banterla, F, Bresin, E, Tripodo, C, Bettoni, S, Donadelli, R, Valoti, E, Tedesco, F, Amore, A, Coppo, R, Ruggenenti, P, Gotti, E, Remuzzi, G: Dynamics of complement activation in aHUS and how to monitor eculizumab therapy. *Blood,* 124**:** 1715-1726, 2014.
9. Pechtl, IC, Kavanagh, D, McIntosh, N, Harris, CL, Barlow, PN: Disease-associated N-terminal complement factor H mutations perturb cofactor and decay-accelerating activities. *J Biol Chem,* 286**:** 11082-11090, 2011.
10. Servais, A, Noel, LH, Roumenina, LT, Le Quintrec, M, Ngo, S, Dragon-Durey, MA, Macher, MA, Zuber, J, Karras, A, Provot, F, Moulin, B, Grunfeld, JP, Niaudet, P, Lesavre, P, Fremeaux-Bacchi, V: Acquired and genetic complement abnormalities play a critical role in dense deposit disease and other C3 glomerulopathies. *Kidney Int,* 82**:** 454-464, 2012.
11. Maga, TK, Nishimura, CJ, Weaver, AE, Frees, KL, Smith, RJ: Mutations in alternative pathway complement proteins in American patients with atypical hemolytic uremic syndrome. *Hum Mutat,* 31**:** E1445-1460, 2010.
12. Caprioli, J, Noris, M, Brioschi, S, Pianetti, G, Castelletti, F, Bettinaglio, P, Mele, C, Bresin, E, Cassis, L, Gamba, S, Porrati, F, Bucchioni, S, Monteferrante, G, Fang, CJ, Liszewski, MK, Kavanagh, D, Atkinson, JP, Remuzzi, G: Genetics of HUS: the impact of MCP, CFH, and IF mutations on clinical presentation, response to treatment, and outcome. *Blood,* 108**:** 1267-1279, 2006.
13. Kavanagh, D, Goodship, TH, Richards, A: Atypical hemolytic uremic syndrome. *Semin Nephrol,* 33**:** 508-530, 2013.
14. Delvaeye, M, Noris, M, De Vriese, A, Esmon, CT, Esmon, NL, Ferrell, G, Del-Favero, J, Plaisance, S, Claes, B, Lambrechts, D, Zoja, C, Remuzzi, G, Conway, EM: Thrombomodulin mutations in atypical hemolytic-uremic syndrome. *N Engl J Med,* 361**:** 345-357, 2009.