SUPPLEMENTARY MATERIAL

TABLE OF CONTENTS

Definitions of clinical outcomes

Clinical remission for membranous nephropathy (MN) was defined according to 2012 Kidney Disease: Improving Global Outcomes (KDIGO clinical practice guideline for glomerulonephritis[1]) as 1) complete in case of urinary protein excretion less than 500 mg per day or 500 mg/g creatinine; 2) partial in case of urinary protein excretion < 3.5 g per day or 3500 mg/g creatinine and $\geq$ 500 mg/g creatinine with at least 50% reduction compared to baseline. Antibody depletion was defined as complete disappearance of antibodies in PLA_2R-Ab positive patients.

Complete-cases datasets and collinear cell subsets

The analyses excluded 5 healthy controls because of missing data, and 39 cell subsets because of redundancy. Out of the original 68 cell subsets, 11 were excluded because they were missing in over 10% of patients. Among the remaining 57 cell subsets, we identified those that were collinear using the Stata built-in command *_rmcoll*. To visualize the strength of their correlation, we used a correlation plot (**Supplementary Figure S1**). Based on the degree of redundancy evaluated by statistical collinearity and correlation as well as by background knowledge, we selected 30 cell subsets (73 subjects) for the main analyses ("main-analyses reduced dataset").

To check for consistency of our findings, we created one additional data set in which we kept most of the collinear cell subsets (52 cell subsets in total, 71 subjects; **Supplementary Table S1**), and used this dataset for a sensitivity analysis; in this setting, we used LASSO (least absolute shrinkage and selection operator). LASSO is meant for high-dimensional models having too many potential covariates (possibly highly correlated) for the sample size at hand. Among the 52 cell subsets, we used LASSO to identify which predicted MN as the dependent dichotomous variable (=1 if MN and 0 if Healthy controls or CKD, pooled) via logistic regression. The LASSO coefficient path (**Supplementary Figure S3**) carried out on 52 cell subsets and 71 subjects, confirmed the main findings obtained from the main-analyses reduced dataset (30 cell subsets, and 73 subjects). An

additional sensitivity analysis was performed by calculating variable importance via random forest on the main-analyses reduced dataset (30 cell subsets, and 73 subjects) (**Supplementary Figure S2**, and see main body of the manuscript for details).

Quantification of anti-PLA_2R antibodies

To quantify anti-PLA_2R Ab levels in the cell supernatants, we used a commercial quantitative ELISA assay (EUROIMMUN AG) for the IgG-specific isotype of anti-PLA_2R, according to manufacturer instructions. Calibrators with anti-PLA_2R concentration of 2, 100, 500, and 1500 RU per ml and positive and negative controls were added to the microplates. Optical density was read at 450 nm using a microplate photometer (Multiskan FC, ThermoFisher Scientific Inc). Serums anti-PLA_2R IgG antibodies were assessed similarly at the same institutions where patients were enrolled.

Analyses of PhIP-seq results

To analyze PhIP-seq results, the counts per million (CPM) matrix indicating the fraction of reads aligning to each phage clone in each sample was calculated using Kallisto.[2] The CPM matrix was corrected for batch effects using the phip-stat (https://github.com/lasersonlab/phip-stat) software with the command “phip clipped-factorization-model --no-normalize-to-reads-per-million -i cpm.tsv -o facotrized.tsv”. Hits were called using the phip-stat command "phip call-hits -i factorized.tsv -o hits.tsv --fdr 0.05 --normalize-to-reads-per-million never --reference-quantile 1.0”. Differences in the total PhIP-seq hit counts between MN, CKD, and HC samples were assessed using Welch’s t-test. To search for enrichments in hits against particular genes (human90) or viruses (virscan) in the MN patients, hits were grouped according to gene, pfam identifiers,[3] and viral family, genus, species, and subspecies. Total hits for each group (e.g. "viral family: Herpesviridae”) were tested for enrichment in MN vs. CKD, MN vs. HC, and MN vs. CKD + HC using Fisher’s exact test, with the false discovery rate controlled at alpha=0.05 using the Benjamini/Hochberg procedure. To visualize overall inter-sample similarity (Jaccard similarity of the hits for each pair of samples), t-distributed

Stochastic Neighbor Embedding[4] embeddings were calculated using the t-SNE routine in scikit-learn.[5]

SUPPLEMENTARY REFERENCES

1. Chapter 7: Idiopathic membranous nephropathy. *Kidney Int Suppl (2011),* 2**:** 186-197, 2012.
2. Bray, NL, Pimentel, H, Melsted, P, Pachter, L: Near-optimal probabilistic RNA-seq quantification. *Nat Biotechnol,* 34**:** 525-527, 2016.
3. Finn, RD, Bateman, A, Clements, J, Coggill, P, Eberhardt, RY, Eddy, SR, Heger, A, Hetherington, K, Holm, L, Mistry, J, Sonnhammer, ELL, Tate, J, Punta, M: Pfam: the protein families database. *Nucleic Acids Res,* 42**:** D222-D230, 2014.
4. van der Maaten, L, Hinton, GE: Visualizing data using t-SNE. *J Mach LearnResearch,* 9, 2008.
5. Pedregosa, F, Varoquaux, G, Gramfort, A, Michel, V, Thirion, B, Grisel, O, Blondel, M, Prettenhofer, P, Weiss, R, Dubourg, V: Scikit-learn: Machine learning in Python. *Journal of machine learning research,* 12**:** 2825-2830, 2011.

Supplementary Table S1. Analyzed cell subsets.

Subset name	Abbreviation	Markers	Parent Population
Total B cells	Bcell	CD3-CD56-CD19+	Singlets
Naïve/transitional B cells	NaiveB	CD3-CD56-CD19+IgD+CD27-	CD3-CD56-CD19+
Unswitched memory B cells	UnsBmem	CD3-CD56-CD19+IgD+CD27+	CD3-CD56-CD19+
Switched memory B cells	SwtBmem	CD3-CD56-CD19+ IgD-CD27+	CD3-CD56-CD19+
Double negative memory B cells	DNBmem	CD3-CD56-CD19+IgD-CD27-	CD3-CD56-CD19+
T3 B cells	T3Bcell	CD3-CD56-CD19+IgD+CD27-CD24lowCD38low	CD3-CD56-CD19+
T1 B cells	T1Bcell	CD3-CD56-CD19+IgD+CD27-CD24+highCD38+high	CD3-CD56-CD19+
T2 B cells	T2Bcell	CD3-CD56-CD19+IgD+CD27-CD24intCD38int	CD3-CD56-CD19+
Switched/resting B cells	SwtBrest	CD3-CD56-CD19+IgD-CD27+CD24+CD95-CD21+	CD3-CD56-CD19+
Switched/activated B cells	SwtBact	CD3-CD56-CD19+IgD-CD27+CD24-CD95+CD21-	CD3-CD56-CD19+
Double negative B cells	CD21+DNBcell	CD3-CD56-CD19+IgD-CD27-CD21+CD95-	CD3-CD56-CD19+
T1+T2 B cells	T1+T2 B cells	CD3-CD56-CD19+IgD+CD27-CD24int/highCD38int/high	CD3-CD56-CD19+
B_{regs}	B_{REG}	CD3-CD56-CD19+CD25+highCD71+high	CD3-CD56-CD19+
Ab secreting B cells_1	AbBcell_1	CD3-CD56-CD19+CD38+CD27+	CD3-CD56-CD19+
Ab secreting B cells_2	AbBcell_2	CD3-CD56-CD19+CD38+highCD27+high	CD3-CD56-CD19+
Plasma cell_1	PC-1	CD3-CD56-CD19+CD38+highCD27+highCD138+	CD3-CD56-CD19+CD38+highCD27+high
Plasmablast_1	PB-1	CD3-CD56-CD19+CD38+highCD27+highCD138-	CD3-CD56-CD19+CD38+highCD27+high
Plasma cell_2	PC-2	CD3-CD56-CD19+CD38+highCD27+highCD138+	CD3-CD56-CD19+
Plasmablast_2	PB-2	CD3-CD56-CD19+CD38+highCD27+highCD138-	CD3-CD56-CD19+
Plasma cell_3	PC-3	CD138-CD38high	Singlets
Plasmablast_3	PB-3	CD138+CD38high	Singlets
Plasma cell_4	PC-4	CD3-CD56-CD19+CD38+highCD27+high	CD3-CD56-CD19+
IL17+ B cells	IL17Bcell	CD19+IL17+	CD19+
IFNγ+ B cells	IFNgBcell	CD19+IFNγ+IL4-	CD19+
IFNγ+IL4+ B cells	IFNgIL4Bcell	CD19+IFNγ+IL4+	CD19+
IL4+ B cells	IL4Bcell	CD19+IFNγ-IL4+	CD19+
Activated lymphocytes	Active_Lymph	CD38high	Singlets
Memory B cells	Bmem	CD3-CD56-CD19+CD21-	CD3-CD56-CD19+

Activated B cells	Bact	CD3-CD56-CD19+IgD-CD27-CD21+CD95-	CD3-CD56-CD19+
Naïve CD4 T cells_1	NaiveCD4-1	CD8-CD4+CD45RA+CD27+	CD8-CD4+
Effector CD4 T cells	CD4Eff	CD8-CD4+CD45RO+CD27-	CD8-CD4+
Central Memory CD4 T cells	CD4Mem	CD8-CD4+CD45RO+CD27+	CD8-CD4+
Effector CD8 T cells	CD8Eff	CD4-CD8+CD28+	CD4-CD8+
Memory CD8 T cells	CD8Mem	CD4-CD8+CD28-	CD4-CD8+
T_{regs}	T_{REG}-1	CD3+CD8-CD4+CD25+CD127low	CD3+CD8-CD4+
Active T_{regs}*_1*	T_{REG}-2	CD3+CD8-CD4+CD25+CD127lowCCR4+CD45RA-	CD3+CD8-CD4+
Total T cells	Tcell	CD3+	Singlets
CD4 T cells	CD4Tcell	CD3+CD8-CD4+	CD3+
Exhausted CD4 T cells_1	CD4Texh-1	CD3+CD8-CD4+PD1+CD57-	CD3+CD8-CD4+
Exhausted CD4 T cells_2	CD4Texh-2	CD3+CD8-CD4+PD1+CD57-KLRG1+	CD3+CD8-CD4+PD1+CD57-
Anergic CD4 T cells_1	CD4Tan-1	CD3+CD8-CD4+PD1+CD57-KLRG1-	CD3+CD8-CD4+PD1+CD57-
Senescent CD4 T cells_1	CD4Tsen-1	CD3+CD8-CD4+PD1-CD57+KLRG1+	CD3+CD8-CD4+PD1-CD57+
CD8 T cells	CD8Tcell	CD3+CD4-CD8+	CD3+
Exhausted CD8 T cells_1	CD8Texh-1	CD3+CD4-CD8+PD1+CD57-	CD3+CD4-CD8+
Exhausted CD8 T cells_2	CD8Texh-2	CD3+CD4-CD8+PD1+CD57-KLRG1+	CD3+CD4-CD8+PD1+CD57-
Anergic CD8 T cells_1	CD8Tan-1	CD3+CD4-CD8+PD1+CD57-KLRG1-	CD3+CD4-CD8+PD1+CD57-
Senescent CD8 T cells_1		CD3+CD4-CD8+PD1+CD57+	CD3+CD4-CD8+PD1-CD57+
Exhausted CD4 T cells_3	CD4Texh-3	CD3+CD8-CD4+PD1+CD57-KLRG+	CD3+CD8-CD4+
Anergic CD4 T cells_2	CD4Tan-2	CD3+CD8-CD4+PD1+CD57-KLRG1-	CD3+CD8-CD4+
Senescent CD4 T cells_2	CD4Tsen-2	CD3+CD8-CD4+PD1-CD57+KLRG1+	CD3+CD8-CD4+
Exhausted CD8 T cells_3	CD8Texh-3	CD3+CD4-CD8+PD1+CD57-KLRG1+	CD3+CD4-CD8+
Anergic CD8 T cells_2	CD8Tan-2	CD3+CD4-CD8+PD1+CD57-KLRG1-	CD3+CD4-CD8+
Senescent CD8 T cells_2	CD8Tsen	CD3+CD4-CD8+PD1-CD57+KLRG1+	CD3+CD4-CD8+
Total TFH cells	Tfh	CD8-CD4+CXCR5+PD1+	CD8-CD4+
TFH17	Tfh17	CD8-CD4+CXCR5+PD1+CCR6+CXCR3-	CD8-CD4+
TFH1	Tfh1	CD8-CD4+CXCR5+PD1+CCR6-CXCR3+	CD8-CD4+
TFH2	Tfh2	CD8-CD4+CXCR5+ PD1+CCR6-CXCR3-	CD8-CD4+
IL17+ CD4 T cells	Th17	CD8-CD4+IL17+	CD8-CD4+
IFNγ+ CD4 T cells	Th1	CD8-CD4+IFNγ+IL4-	CD8-CD4+
IFNγ+IL4+ CD4 T cells	IFNgIL4CD4	CD8-CD4+IFNγ+IL4+	CD8-CD4+
IL4+ CD4 T cells	Th2	CD8-CD4+IFNγ-IL4+	CD8-CD4+
IL17+ CD8 T cells	IL17CD8	CD4-CD8+IL17+	CD4-CD8+

IFNγ+ CD8 T cells	IFNgBD8	CD4-CD8+IFNγ+IL4-	CD4-CD8+
IFNγ+IL4+ CD8 T cells	IFNgIL4CD8	CD4-CD8+IFNγ+IL4+	CD4-CD8+
IL4+ CD8 T cells	IL4CD8	CD4-CD8+IFNγ-IL4+	CD4-CD8+
Active T_{regs}_2	T_{REG}-3	CD3+CD8-CD4+CCR4+CD45RA-CD25+CD127low	CD3+CD8-CD4+CCR4+CD45RA-
Naïve CD4 T cells_2	NaiveCD4-2	CD8-CD4+CD45RA+CD45RO-CD27+	CD8-CD4+
Naïve CD8 T cells	NaiveCD8	CD4-CD8+CD45RA+CD45RO -CD27+	CD4-CD8+

Supplementary Table S2. Analyzed cytokines and chemokines.

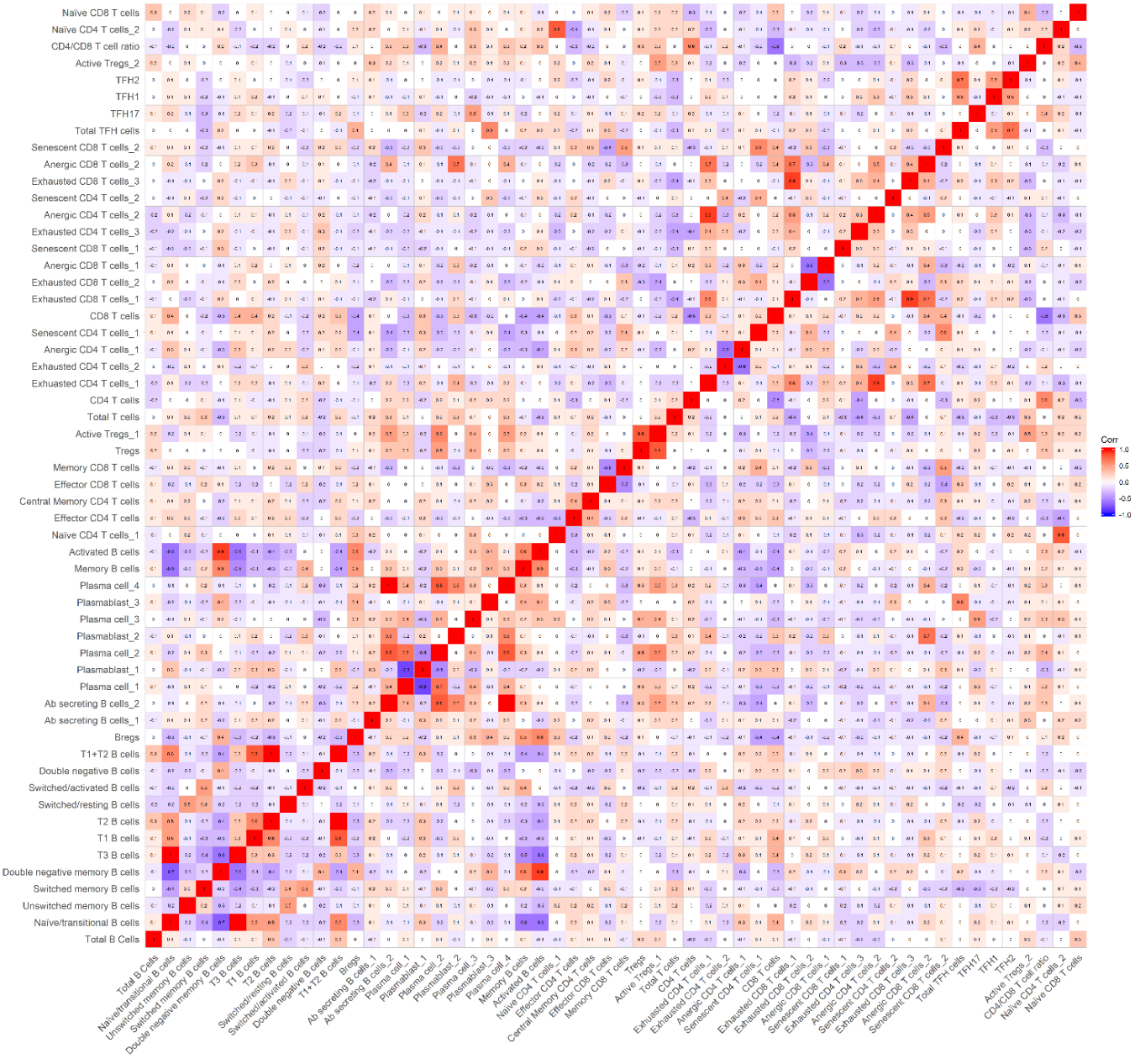
Cytokine/Chemokine (pg/ml)	MN (n = 19)	HC (n = 8)
CD40L	ND	ND
G-CSF/CSF-3	ND	ND
GM-CSF	ND	ND
IFN-g	ND	ND
IL-1β	ND	ND
IL-10	ND	ND
IL-12p70	ND	ND
IL-13	ND	ND
IL-17A	ND	ND
IL-2	ND	ND
IL-4	ND	ND
IL-5	ND	ND
IL-6	ND	ND
IL-7	0.64 ± 0.40	0.31 ± 0.28
IL-8	ND	ND
MCP-1	22.10 ± 13.16	21.45 ± 7.74
MIP-1β	ND	ND
TNF-α	0.03 ± 0.01*	0.02 ± 0.00
TNF-β	ND	ND

ND = not detectable in over 90% of samples. *P<0.05.

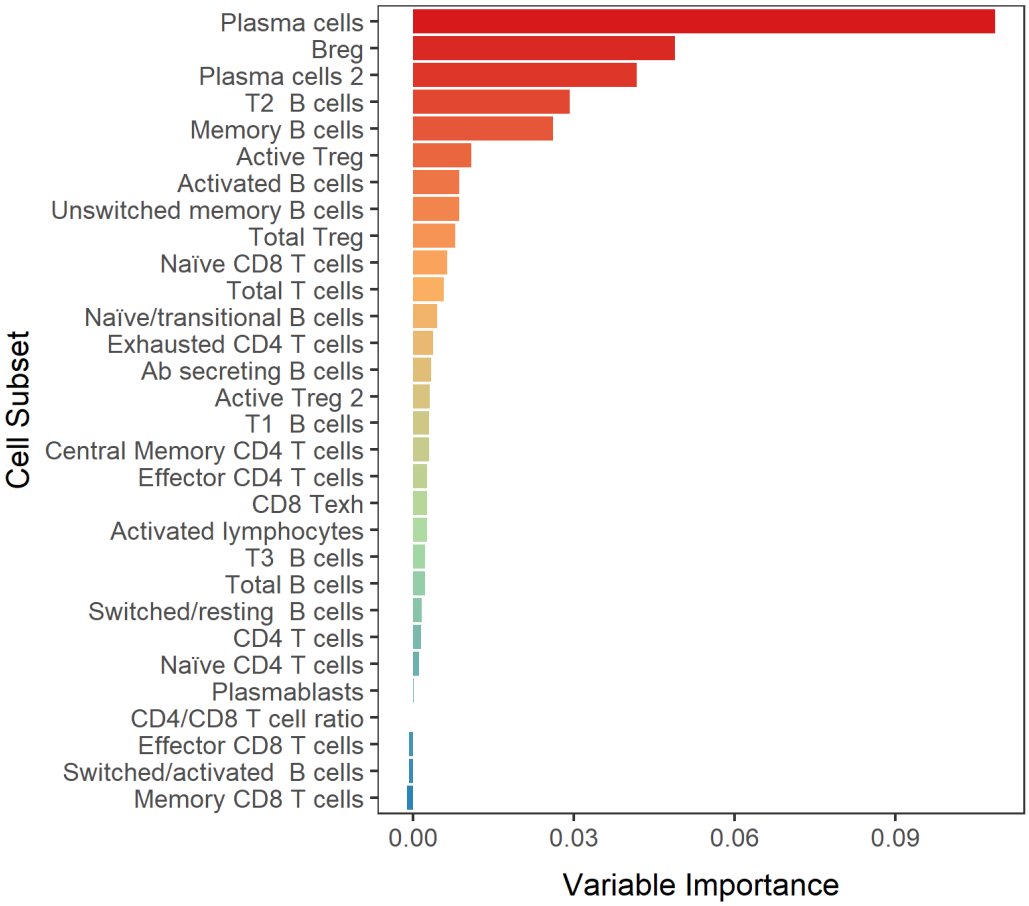
Supplementary Table S3. PhIP-seq results. Detected antibody specificities for each sample (1=hit detected, 0=no hit detected) with association tests for enrichment in the MN, CKD, or HC groups (first sheet). Also shown are association tests for hits grouped by gene, protein, or viral taxon (second sheet). Association tests were performed using Fisher's exact test and the false discovery rate was controlled using the Benjamini/Hochberg procedure (see **Supplemental spreadsheet**).

Supplementary information is available at KI Report's website.

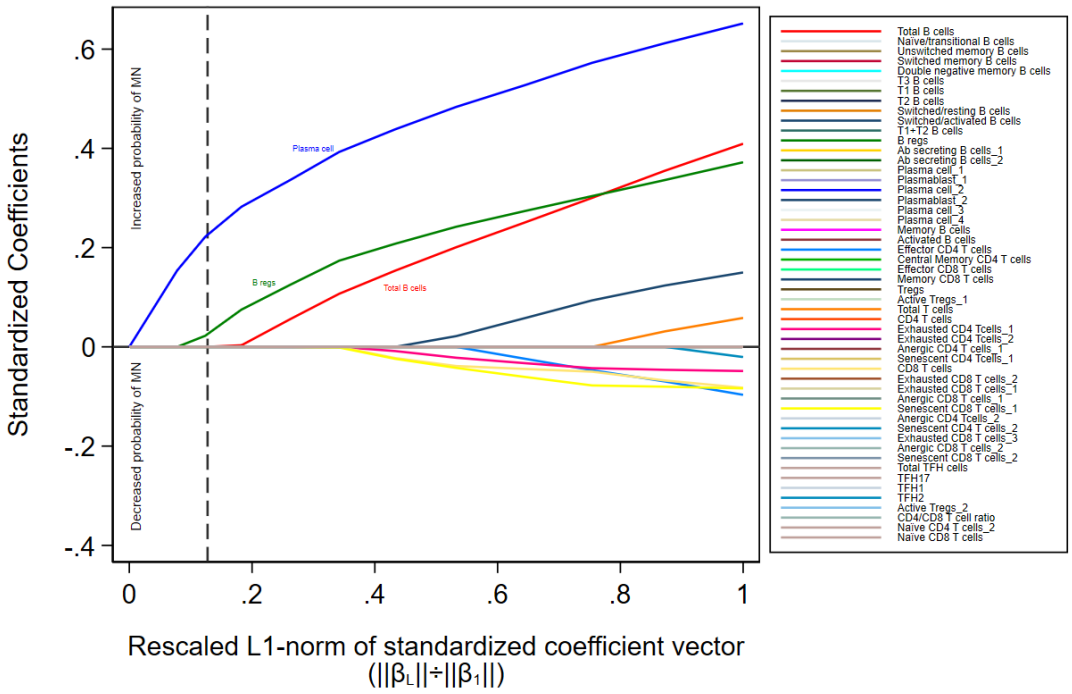
SUPPLEMENTAL FIGURES



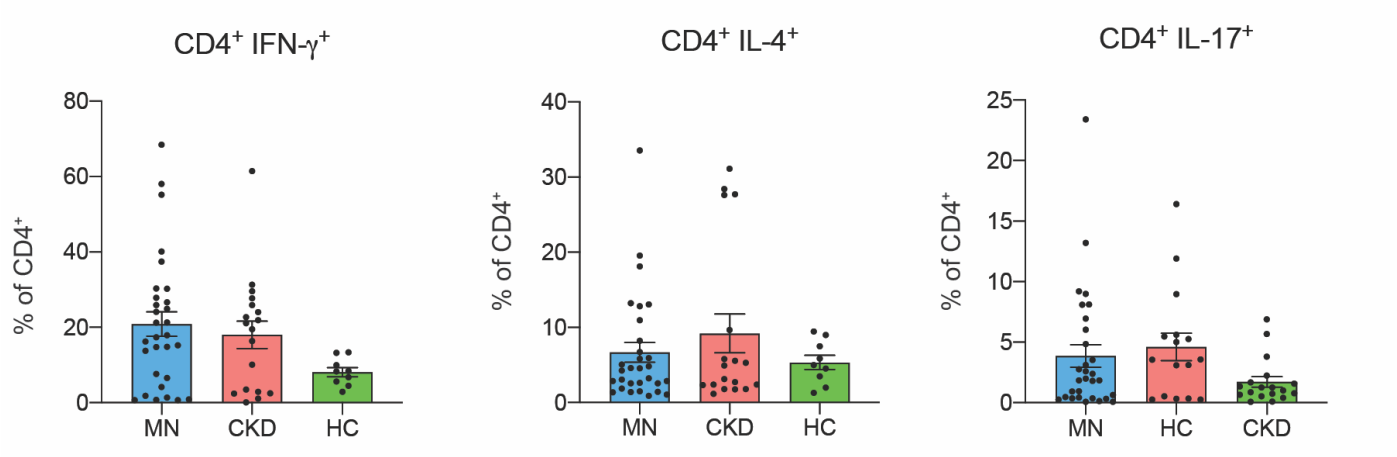
Supplementary Figure S1. Correlation plot of the 52 cell subsets in 71 patients with complete data (30 MN, 31 CKD, 10 HC).



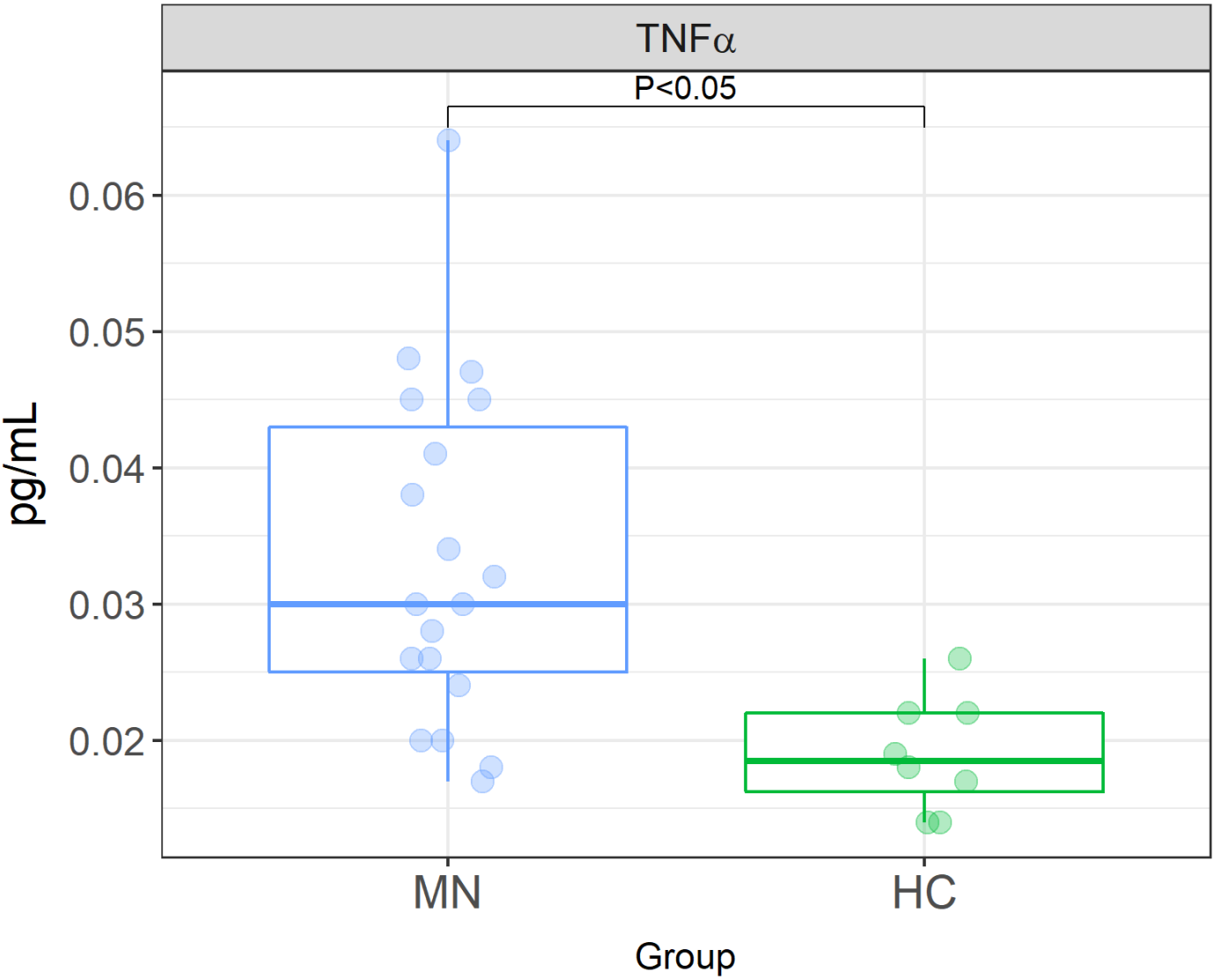
Supplementary Figure S2. Variable importance from random forest analysis. A large Variable IMPortance (VIMP) value for a cell subset indicates that misspecification of that variable decreases ability in the forest in distinguishing between subjects with membranous nephropathy and the other two conditions, based on the cell subsets. VIMP close to zero indicates that the cell subset contributes nothing to distinguishing between groups, and negative values indicate that the ability to distinguishing between groups improves when the cell subset is misdefined. In the latter case, we assume noise is more informative than the true cell subset. As such, if we ignore cell subsets with negative and near zero values of VIMP, we are left mostly with B cell subsets. Among them, plasma cells and B_{REG} had the largest VIMP.



Supplementary Figure S3. LASSO for variable selection. Standardized coefficient estimates as a function of the tuning parameter which “shrinks" the coefficient towards zero as its value (on x-axis) gets larger; by setting some coefficient to zero, the tuning parameter determines which variables the LASSO will eventually exclude. Plasma cells and B_{REG} had the largest standardized coefficients and would have been selected (along with B cells) as the only significant cell subsets had the tuning parameter been set to the level of the vertical dotted line.



Supplementary Figure S4. IFN-g$^+$, IL-4$^+$, and IL-17$^+$ CD4$^+$ T cell subsets in MN, CKD patients and healthy controls (HC). Data are presented as mean ± SD. Each dot represents the value for a single patient.



Supplementary Figure S5. Serum levels of TNF-α. Serum levels of TNF-α in patients with membranous nephropathy (MN) and in healthy controls (HC). Data are presented as mean ± SD. Each dot represents the value for a single patient. *P*<0.05 by unpaired T-test and Holm-Bonferroni adjustment for multiple comparisons.