

## Molecular Stratification of Chronic Kidney Disease

### Authors:

Anna Reznichenko<sup>1\*</sup>, Viji Nair<sup>2</sup>, Sean Eddy<sup>2</sup>, Mark Tomilo<sup>2</sup>, Timothy Slidel<sup>3</sup>, Wenjun Ju<sup>2</sup>, James P. Conway<sup>4</sup>, Shawn S. Badal<sup>5</sup>, Johnna Wesley<sup>6</sup>, John T. Liles<sup>5</sup>, Sven Moosmang<sup>1</sup>, Julie M. Williams<sup>7</sup>, Carol Moreno Quinn<sup>8</sup>, Markus Bitzer<sup>2</sup>, Anil Karihaloo<sup>6</sup>, Matthew D. Breyer<sup>9</sup>, Kevin L. Duffin<sup>10</sup>, Uptal D. Patel<sup>3</sup>, Maria Chiara Magnone<sup>9</sup>, Ratan Bhat<sup>11</sup>, Matthias Kretzler<sup>2\*</sup>

### Affiliations:

<sup>1</sup>Translational Science & Experimental Medicine, Research and Early Development, Cardiovascular, Renal and Metabolism, BioPharmaceuticals R&D, AstraZeneca, Gothenburg, Sweden.

<sup>2</sup>Michigan Medicine, Ann Arbor, MI, USA.

<sup>3</sup>Early Computational Oncology, Translational Medicine, Research and Early Development, Oncology R&D, AstraZeneca, Cambridge, UK.

<sup>4</sup>Early Computational Oncology, Translational Medicine, Research and Early Development, Oncology R&D, AstraZeneca, Gaithersburg, USA.

<sup>5</sup>Gilead Sciences, Foster City, CA, USA.

<sup>6</sup>Global Diabetes, Cardio-renal and Translational Research, NovoNordisk, Seattle, WA, USA.

<sup>7</sup>Bioscience Renal, Research and Early Development, Cardiovascular, Renal & Metabolism, BioPharmaceuticals R&D, AstraZeneca, Gothenburg, Sweden.

<sup>8</sup>Medical Affairs, Vifor Pharma, Zurich, Switzerland.

<sup>9</sup>Cardiovascular and Metabolism Therapeutic Area, Janssen Research and Development LLC, Boston, MA, USA

<sup>10</sup>Diabetes, Eli Lilly and Company, Indianapolis, IN, USA.

<sup>11</sup>Search and Evaluation, Cardiovascular Renal & Metabolism, Business Development & Licensing, BioPharmaceuticals R&D, AstraZeneca, Gothenburg, Sweden.

### \*Corresponding authors:

Anna Reznichenko, M.D., Ph.D.  
Principal Scientist  
Translational Science & Experimental  
Medicine  
Research and Early Development,  
Cardiovascular, Renal and Metabolism,  
BioPharmaceuticals R&D,  
AstraZeneca, Gothenburg, Sweden  
Pepparedsleden 1,  
SE-431 83 Mölndal, Sweden  
Tel +46 (0)72 515 82 70  
[anna.reznichenko@astrazeneca.com](mailto:anna.reznichenko@astrazeneca.com)

MATTHIAS KRETZLER, M.D.  
Warner-Lambert/Parke-Davis Professor of  
Medicine  
Nephrology/Internal Medicine and  
Computational Medicine and Bioinformatics  
University of Michigan  
MSRB II, 4544-D  
1150 W. Medical Center Dr.  
Ann Arbor, MI 48109  
734-615-5757  
fax: 734-763-0982  
[kretzler@umich.edu](mailto:kretzler@umich.edu)

## **One sentence summary:**

Unbiased grouping of patients based on kidney biopsy transcriptomics profiles generated a novel molecular categorization of chronic kidney disease.

## **Abstract**

Current classification of chronic kidney disease (CKD) into stages based on the indirect measures of kidney functional state, estimated glomerular filtration rate and albuminuria, is agnostic to the heterogeneity of underlying etiologies, histopathology, and molecular processes. We used genome-wide transcriptomics from patients' kidney biopsies, directly reflecting kidney biological processes, to stratify patients from three independent CKD cohorts. Unsupervised Self-Organizing Maps (SOM), an artificial neural network algorithm, assembled CKD patients into four novel subgroups, molecular categories, based on the similarity of their kidney transcriptomics profiles. The unbiased, molecular categories were present across CKD stages and histopathological diagnoses, highlighting heterogeneity of conventional clinical subgroups at the molecular level. CKD molecular categories were distinct in terms of biological pathways, transcriptional regulation and associated kidney cell types, indicating that the molecular categorization is founded on biologically meaningful mechanisms. Importantly, our results revealed that not all biological pathways are equally activated in all patients; instead, different pathways could be more dominant in different subgroups and thereby differentially influencing disease progression and outcomes. This first kidney-centric unbiased categorization of CKD paves the way to an integrated clinical, morphological and molecular diagnosis. This is a key step towards enabling precision medicine for this heterogeneous condition with the potential to advance biological understanding, clinical management, and drug development, as well as establish a roadmap for molecular reclassification of CKD and other complex diseases.

## Introduction

Chronic kidney disease (CKD) is a global health burden affecting nearly 700 million people worldwide, with a high cost of care to individuals and health systems. It is an independent risk factor for cardiovascular disease and is associated with increased morbidity and mortality (1, 2).

Biologically, CKD is a highly heterogeneous group of disorders characterized by sustained alterations in kidney structure and function arising from a wide range of etiologies (the most common are diabetic and hypertensive nephropathies) and associated with a multitude of underlying molecular processes in the kidney (3). Even within the same etiological group, individual disease presentation, histopathological features, progression rates and treatment responses are variable, reflecting the underlying biological heterogeneity.

According to the KDIGO (Kidney Disease Improving Global Outcomes) guidelines, CKD is classified into five stages (CKD 1-5) based on gradations of glomerular filtration rate (GFR), estimated from serum creatinine, and/or albuminuria (3). This, now widely accepted classification system, was a major achievement in the field as it provided a common vocabulary and standardized approach to disease management (4-7). However, the classification relies on systemic, indirect measures and is therefore agnostic to the intra-renal biology. As a result, biologically diverse cases presenting with the same GFR or albuminuria values are classified under the same CKD stage category, thereby precluding personalized prognosis and treatment options (8).

Histopathological diagnosis based on kidney biopsy provides more direct insights on the intra-renal state (9). However, several tissue morphological features may overlap between various kidney diseases. Furthermore, tissue molecular signatures also overlap, as previously revealed by a transcriptomics study of CKD biopsies, where patterns of molecular similarity spanned histopathological diagnoses (10).

Thus, an unbiased re-categorization of CKD by using kidney molecular profiles may circumvent the shortcomings of GFR staging or histopathological grouping. In a field where specific therapies are still lacking, such a mechanistic classification based on underlying disease biology is urgently needed to fuel diagnostic, prognostic, and therapeutic development that will enable the transition to precision medicine (11-13).

The overarching goal of this study was to determine whether a mechanistic disease classification based on intra-renal molecular characteristics can be achieved across multiple cohorts, as a critical step towards development of novel personalized treatments for CKD. We used individual transcriptomics profiles from kidney biopsies to group patients in an unbiased, data-driven manner. We analyzed an extensive collection of kidney transcriptomics data across three CKD patient cohorts from North America and Europe that encompass a broad range of kidney function and wide spectrum of diseases. Focusing on the tubulointerstitial compartment, the most abundantly available tissue fraction in kidney biopsies and closely linked to long-term outcomes of CKD (40), we employed an unsupervised artificial neural network algorithm, Self-Organizing Maps (SOM) (14, 15), to reveal inherent patient sub-groups. We then substantiated the clinical relevance of these sub-groups by assessing their association with biological pathways and clinical parameters.

## Results

### ***CKD patient population map construction and segmentation***

A total of 314 CKD patients from European Renal cDNA Bank (ERCB)(16), Clinical Phenotyping Resource and Biobank Core (C-PROBE)(17), and Nephrotic Syndrome Study Network (NEPTUNE)(18) cohorts with available genome-wide transcriptomics data from kidney biopsy tubulointerstitial fraction were included in this study. Patient clinical characteristics are summarized in *Table 1*.

Based on the large sample size as well as broadest range of kidney function and spectrum of histopathological diagnoses, the ERCB dataset was selected as the discovery dataset to maximize the pattern detection potential in the data. The independent datasets from C-PROBE and NEPTUNE, contributing complementary GFR ranges and etiologies, were then used for validation.

An unsupervised SOM of the ERCB CKD population was constructed based solely on patient kidney transcriptomics profiles, comprised of expression values for 8,454 genes. The algorithm mapped patients onto the 8x8 SOM grid, resulting in 1 to 8 individuals (median of 2) per each of the 64 SOM units (*Fig 1A*), placed in a topological order based on the relative similarities of their transcriptional profiles. Hierarchical clustering of SOM units was then performed to delineate subgroups with similar gene expression profiles. Four inherent patient clusters were thus identified within the discovery CKD population, named “molecular categories” and assigned arbitrary colors – Blue, Gold, Olive, Plum – in line with the unbiased nature of their discovery (*Fig 1B*). The larger Blue and Plum molecular categories accounted for 35% and 27% of the cases while 20% and 18% of the patients mapped to the Gold and Olive categories, respectively. Global gene expression levels were not different between the four molecular categories (*Fig S1*), indicating absence of technical bias but rather, qualitative differences in

patterns underlying the subgrouping. This was further evident from inspecting the transcriptomics profiles visualization, showing distinct patterns between the molecular categories (*Fig 1C*).

### ***CKD stages consist of molecularly heterogeneous categories***

Having thus identified different molecular categories of patients with CKD based on kidney transcriptomics features, we next wanted to compare them with clinical CKD classifications.

*Figs 1D-E* visualize the correspondence between novel molecular categories and conventional clinical classification. The molecular categories were present across all CKD stages and histopathological diagnoses, highlighting heterogeneity of clinical subgroups at the molecular level. Gold and Plum categories were relatively more frequent in advanced disease stages (CKD 3-5), while CKD 1-3 patients were predominantly mapped to Blue and Olive categories (*Fig 1D*). However, the fact that the same molecular categories were observed at different CKD stages indicates independence from GFR as a potential confounder (*Fig S2* shows similar estimated GFR levels across molecular categories within same CKD stages strata). Likewise, each histopathologically defined patient group (*Fig 1E*) contained multiple molecular categories, highlighting within-group heterogeneity as well as between-group similarities. *Fig 2* illustrates the kidney molecular heterogeneity in a sample of 9 patients (three patients per each of three CKD histopathological groups: focal segmental glomerulosclerosis, diabetic nephropathy, and immunoglobulin A nephropathy) and highlights differences in grouping in the two classification approaches.

Therefore, CKD molecular categories represent a novel characterization, not captured by existing classifications, and provide orthogonal biomechanistic information, independent of clinical manifestations or disease severity.

### ***CKD molecular categories are biologically distinct***

In order to obtain functional insights into biological processes underlying the molecular categorization, we compared differences in kidney gene expression between each molecular category to those of healthy controls (living kidney donors). The volcano plots (*Fig 3 A*), highlight substantial transcriptomics changes in each CKD molecular categories. Interestingly, the numbers of significantly differentially expressed genes (DEGs) in each comparison were not driven by the differences in sample size. Blue, the largest CKD molecular category, yielded the fewest number of differentially expressed genes (216), predominantly down-regulated. In contrast, Plum and Gold categories had many more DEGs (3,125 and 3,472, respectively), evenly proportioned between up- and down-regulation (*Table 2*). Finally, when the CKD samples were pooled together and contrasted to the healthy controls in a conventional case-control analysis, the differential expression was less pronounced (1,341 DEGs), likely dampened by the heterogeneity of cases. The list of all significant DEGs (at FDR  $q < 0.0001$ ) per CKD molecular category is provided in *Table S1*.

Specific biological themes or pathways in each CKD molecular category were then identified using gene set enrichment analysis (GSEA). In the first instance, a hypothesis-free GSEA using canonical, CKD-agnostic gene sets (Hallmark, Curated Canonical Pathways, Gene Ontology) was performed. The resulting heatmap of gene set enrichment score values (*Fig 3 B*) shows distinct patterns across the molecular categories and presence of both unique and shared pathways, as corroborated by Venn diagrams of the significantly enriched (FDR  $q < 0.05$ ) up- and down-regulated gene sets (*Fig 3 C and D*, respectively). The top 10 uniquely enriched gene sets (*Table 2*, while a complete list of significantly enriched gene sets per CKD molecular category is provided in *Table S2*), indicate pathways related to transcription, signaling, response to stimuli, apoptosis were selectively down-regulated in the Blue category patients. Gold category revealed a strong pro-inflammatory and pro-fibrotic drive, as the top activated processes

including immune cell proliferation and migration, cytokine production, and extracellular matrix organization, while several metabolic pathways were down-regulated. In the Olive category, mitochondrial and peroxisomal processes were up-regulated while immune and membrane transport pathways were inhibited. The Plum category showed enrichment for signaling, protein synthesis, and vesicle transport processes; at the same time, ion transport pathways were down-regulated.

In line with the vast transcriptomics differences, upstream transcriptional regulator analysis revealed a diverse array of activated regulatory molecules in each of the CKD molecular category (top 20 upstream regulators are shown in *Fig 3 E*, while a complete list per CKD molecular category is provided in *Table S3*).

As the next step, hypothesis-driven gene sets of CKD relevance, including kidney-specific genes, Mendelian kidney disease genes, kidney cell type-specific markers, and kidney injury markers, were tested for enrichment across CKD molecular categories (*Fig 3 F*). All three groups of kidney-specific transcriptome were positively enriched in the Olive category.

Mendelian genes known to cause glomerulopathies but not tubulopathies were enriched in the Olive and Plum categories. The Gold and Plum molecular categories showed enrichment in markers of immune cells and collecting duct, while the Olive category was specifically enriched for the proximal and distal tubular signatures. Furthermore, a set of genes encoding established urinary biomarkers of kidney damage showed diverse trends across the molecular categories (*Fig 3 G*).

Taken together, these results highlight that patients from different CKD molecular categories are distinct in terms of the biological pathways affected in the kidney, indicating that the molecular categorization has a biomechanistic basis.



### ***CKD molecular categories validation in independent cohorts***

We validated our principal findings and determined whether these new molecular categories could be identified in other CKD populations using data from two independent cohorts.

CKD patients from the C-PROBE and NEPTUNE cohorts were mapped onto the discovery SOM based on their kidney biopsy transcriptomics profiles (*Fig 4 A-B*). As a negative control, permuted data produced nonsensical mapping. All four molecular categories were present among the C-PROBE and NEPTUNE patients. The proportions of patients per molecular categories differed between the studied CKD cohorts (chi squared  $p=0.0019$ ; *Fig 4 C*).

To determine whether the underlying differences in biological mechanisms between the CKD molecular categories would manifest in disparate disease progression rates, we analyzed kidney outcomes in patients with available long-term clinical follow-up data (42 C-PROBE and 90 NEPTUNE patients, respectively).

In the C-PROBE cohort, during the follow-up period (median duration 4.7 years, maximal 9.1 years), 7 out of 42 patients progressed to end-stage kidney disease (ESKD), defined as initiation of renal replacement therapy (dialysis or kidney transplantation). Even in this limited analysis, the rate of reaching the end-point varied significantly depending on the molecular category (log-rank  $p=0.028$ ). Kaplan-Meier survival curves stratified by molecular category (*Fig 4D*) demonstrate the significantly higher probability of ESKD incidence in the Gold category, with 5 out of 7 events occurring in this subgroup of patients. The rate of disease progression was thus faster for patients in the Gold category with a median survival probability of 6.8 years, while no other molecular category reached median survival during the follow-up period. Cox regression hazard ratios (HR) for ESKD in Gold compared to other molecular categories was 7.9 [95%CI 1.5-41.1] ( $p=0.014$ ) in the crude model and 5.6 [95%CI 1.0-31.3] in the model

adjusted for baseline eGFR ( $p=0.048$ ). For 37 C-PROBE patients, repeated kidney function values were available over time from which eGFR slopes were calculated. Consistent with the hard end-point (ESKD) prediction, the eGFR slopes also showed a trend for greater annual loss of kidney function in the Gold category (*Fig 4E*).

In NEPTUNE, where the outcome was defined as a composite of ESKD or loss of  $>40\%$  eGFR, there were 16 events among 90 patients during the follow-up period (median duration 4.1 years, maximal 6.0 years). Likewise, the rate of disease progression was significantly different between the molecular categories (log-rank  $p=0.00049$ , *Fig 4F*), with the Gold category patients showing higher event incidence (9 out of 16) and median renal survival probability of 4.5 years. Cox regression HR for Gold compared to other molecular categories was 6.7 [95%CI 2.4-19.0] ( $p=0.00032$ ) in the crude model and 4.3 [95%CI 1.3-14.2] ( $p=0.017$ ) in the model adjusted for baseline eGFR (4 observations were excluded from the adjusted analysis due to missing values).

Interestingly, in contrast to the molecular categories, if stratified by CKD stage at baseline in the C-PROBE cohort (*Fig S3A*) or histopathological diagnosis in NEPTUNE (*Fig S3B*), the event rates were not significantly different.

The presence of the molecular categories in two independent cohorts thus strengthens the generalizability of our findings. The differences in disease progression rates between CKD molecular classes were consistent across both cohorts.

## Discussion

Our findings represent the first kidney transcriptomics-driven approach for molecular categorization of CKD, a significant step towards precision medicine for this complex and heterogeneous condition. This has the potential to transform the CKD field with respect to biological understanding, clinical management, and drug development, as well as pave the way towards molecular reclassification of other complex diseases.

This study was motivated by the growing realization in the field that mechanistic categorization of the various CKD etiologies is lacking (8, 10, 19, 20). The call to redefine CKD to boost therapeutic development is becoming increasingly urgent as its global prevalence is reaching epidemic proportions (21). Integration of molecular features into clinico-morphological diagnosis promises a more mechanistic framework for classification of CKD patients, as a foundation for development of personalized treatment options based on individual disease biology (22-24).

Additive value of molecular characterization on top of clinical parameters has been shown, previously, in studies evaluating the use of circulating or urinary biomarkers to improve disease stratification and prognosis prediction (25, 26). However, a fundamental limitation of such systemic phenotyping is that due to the indirect assessment, the underlying molecular changes in the kidney remain unknown. We pursued a different approach, leveraging available molecular profiles from kidney biopsies, directly reflecting intra-renal biology, as the basis of our categorization strategy.

Transcriptomics profiling provides a fingerprint of genome-wide gene expression patterns reflecting tissue structural and functional states (27-30). These include cell type composition, activity of biological pathways causing, perpetuating or counteracting the disease, interplay of co-morbidities, as well as the effects of treatments (29-32). Our findings highlight the value of

transcriptomics in grouping patients based on similar profiles. The fact that permuted data, where the expression values of each gene were randomized across the samples, failed to map onto SOM signifies the importance of the individual specificity of each patient's transcriptomics profile as well as gene covariance modular structure.

We identified four novel patient subgroups within a CKD population in an unbiased data-driven manner, based on kidney transcriptomics data and without imposing any patient clinical information or preconceived, domain biological knowledge. The emergent molecular categories did not overlap with the conventional clinical classification and were detected orthogonally at all the different CKD stages and histopathological diagnoses, highlighting molecular heterogeneity among CKD patients and underscoring the biological complexity of this multifaceted condition. We propose that molecular information can, therefore, add a new dimension to the existing framework of CKD classification.

Our results highlighted that CKD molecular categories were distinct from the healthy controls as well as one another in terms of biological pathways affected in the kidney. This shows that molecular categorization represents a mechanistic classification of disease that is based on underlying biological processes. Differential expression analysis of the molecular categories vs healthy control produced a multitude of modulated genes. Living kidney donor biopsies were chosen to best represent the "healthy" kidney state, as these individuals were specifically screened for absence of disease and selected for superior kidney function to qualify for donation. Using such a control group, as opposed to commonly used tumor nephrectomy or autopsy tissues, provides a robust baseline and enhances our ability to detect biologically meaningful differences by maximizing the contrast with the disease state. The case-control contrast was further boosted by CKD molecular categories forming intrinsically more uniform subgroups as compared to the heterogeneous all-comer CKD population. This highlights that

the molecular categorization is also able to capture elements of the biological basis of the disease.

Through both hypothesis-free exploration and hypothesis-driven analyses, we found that distinct biological pathways, transcriptional regulators, injury markers, and cell type context were perturbed in each of the CKD molecular categories – an important step in unraveling the biological basis of disease heterogeneity. Many of the enriched pathways, such as inflammation, apoptosis, metabolism, have been previously implicated in CKD (33-36).

However, the novelty of our results is in that not all pathways were found equally activated in all patients; instead, different pathways could be more dominant in different subgroups and thereby differentially influencing disease progression and outcomes. For example, inflammation and fibrosis, generally accepted as universal pathophysiological mechanisms in CKD, showed uneven distribution within the patient population with specific enrichment in one of the molecular categories, CKD-Gold. This challenges existing dogma with implications for disease understanding as well as anti-inflammatory drug development, further emphasizing the importance of precision medicine and mechanistic enrichment approaches to patient selection.

The identified molecular categories were replicated in independent cohorts indicating that the molecular categorization can be extended to other CKD populations. The ability to predict the molecular category of previously unseen patient data strengthens generalizability and robustness of our findings as it demonstrates robustness across CKD etiology and transcriptomics platforms between research centers in the US and Europe. The outcome association was replicated in two independent cohorts. The observation of different ESKD risk across molecular categories conserved between both cohorts is encouraging and advocates for further investigation as the molecular categorization might help identify patients at greater risk for rapid disease progression.

Cross-sectional study design does not allow to observe the anticipated changes over time in the transcriptomics profile following the disease progression processes. However, as repeated kidney biopsies are rarely performed in clinical practice, the temporal trends of molecular changes will need to be ascertained through non-invasive means or through time-course studies in animal models. Finally, to enable implementation in clinical practice, non-invasive biomarkers, e.g. urinary proteins reflecting the intra-renal molecular characteristics of each category, will be needed to allow comprehensive capture in large scale cohorts. Our findings showing differential expression patterns of genes encoding known markers of kidney injury across molecular categories offer a starting point to develop a non-invasive biomarker panel to recognize these categories.

There are several limitations to this study. We analyzed the largest available kidney transcriptomics datasets from CKD patients nevertheless were limited in sample size. Future collaborative effort will be needed to generate more data that may help further refine the classification as well as expand the repertoire of 'omics' beyond transcriptomics. Studied cohorts were subject to an inherent selection bias as patients undergoing clinically indicated kidney biopsy, therefore may differ from the general CKD population in composition. Our analysis focused on the more readily available tubulointerstitial tissue component; we plan to extend similar analysis to the glomerular compartment in the future. However, tubulointerstitial biology is highly relevant for CKD. Tubulointerstitial lesions are known to be a strong predictor of CKD progression. Even in the cases of primary or secondary glomerular disease, studies have suggested that it is the extent of accompanying tubulointerstitial histologic injury that correlates best with renal function decline and progression to ESKD (38-41). Transcriptomics profiling of renal tubulointerstitium has proved informative in deciphering molecular basis of disease (42-44). In the subsequent studies, it would be highly relevant to include also genetics and

comorbidities information for further exploration and interpretation of the characteristics of each molecular class, which for ethical concerns were not available for the discovery cohort.

Our results have the potential to take the CKD disease classification to the next, mechanistic level. A transition to an integrated clinical, morphological and molecular diagnosis is warranted for development of individualized treatments in CKD. Elucidation of the molecular drivers of different patient subgroups can lead to new biological hypotheses, therapeutic targets, and biomarkers. Establishing a first framework of molecular categories in CKD is an important milestone in CKD research, but replication across further disease cohorts and development of non-invasive means to capture the intra-renal categories are urgently needed to accelerate progress towards precision medicine for CKD.

## **Materials & Methods**

### ***Renal transcriptomics data***

We analyzed publicly available renal tubulointerstitial microarray transcriptomics data from three CKD cohorts. Consent was obtained from individual patients at enrollment, and the studies were approved by Institutional Review Boards of participating institutions.

ERCB (European Renal cDNA Bank) is a European multicenter CKD study established to collect renal biopsy tissue for gene expression analysis at the time of a clinically indicated biopsy (45). Affymetrix U133 array transcriptomics data from 165 patients were used for the discovery analysis (GEO accession ID: GSE104954).

C-PROBE (Clinical Phenotyping Resource and Biobank Core) is a multicenter longitudinal observational CKD cohort (17). Tubulointerstitial transcriptomics data generated on Affymetrix Human Genome U133 Plus 2.0 Array were available for 42 patients (GEO accession ID: GSE69438).

NEPTUNE (Nephrotic Syndrome Study Network) is a multi-center, prospective study of children and adults with nephrotic range proteinuria (>500mg/day), recruited at the time of first clinically indicated baseline renal biopsy (46). Affymetrix Human Gene 2.1 ST Array tubulointerstitial transcriptomics data from 107 NEPTUNE-NS adult CKD cases were included in the analysis (GEO accession ID: GSE108112).

Tubulointerstitial transcriptomics data from 46 living kidney donors (GSE104954) were used as controls for differential expression analysis.



### **Self-Organizing Maps algorithm**

All analyses and visualizations were run using R version 3.6.0, unless indicated otherwise. SOM functions implemented in R packages *kohonen* v3.0.8, *popsom* v4.2.1, and *oposSOM* v2.2.0 (49) were used. Inherent data clustering tendency in ERCB dataset was assessed with Hopkins statistics and VAT (Visual Assessment of cluster Tendency) function using *factoextra* R package (Fig. S4). Genes with low variance across samples (median absolute deviation <0.15, 27% of genes) were excluded from the analysis as non-informative. The retained 8,454 gene expression variables were centered and scaled to give them equal importance during the SOM training process and prevent bias from highly abundant genes. Different map grid sizes (4x4 through 12x12) and number of iterations (100 through 50,000) were tested, and mapping quality indices (quantization error, topographic accuracy, embedding accuracy) assessed (Fig. S5). Considering the random nature of SOM initialization, behaviors of SOMs from 10 independent runs were compared (Fig. S5 B). Final model parameters were set to achieve the highest convergence. The optimal map size was determined to be 8x8 (64 SOM units), consistent with the commonly used empirical rule of thumb  $5x\sqrt{N}$  samples (50). Hexagonal topology and a bubble neighborhood function with sum-of-squares as distance measure were used. In order to isolate groups of samples with similar profiles, Ward's hierarchical clustering was run on the codebook vectors. The optimal number of clusters was guided by the 'elbow method' using *NbClust* R package based on the simplest model with the lowest within-cluster sum-of-squares. Diagnostic plots are presented in Fig. S5-S6. Validation mapping of the independent CKD cohorts was performed using 'map.kohonen' function of *kohonen* R package (14). New data for the 8,454 genes were extracted from C-PROBE and NEPTUNE transcriptomics, expression values for missing genes (378 and 427 for C-PROBE and NEPTUNE, respectively) were set to "NA". The data were scaled using the ERCB scaling parameters and then presented to the trained SOM, and distances from the new data points to the closest units were calculated.

## **Downstream analyses**

Individual and group-level transcriptomics profiles were visualized using *oposSOM* R package with default parameters (49). Differential gene expression analyses comparing patients' and control samples were performed with *limma* R package(52). The output was ranked by  $sign(logFC) \cdot -log_{10}(adj.p-value)$ , and functional enrichment analyses were run with Gene Set Enrichment Analysis application using 5,205 gene sets belonging to Hallmark, Curated Canonical Pathways, and Gene Ontology from MSigDB database of molecular signatures (53, 54). Upstream regulator analysis was performed using IPA (QIAGEN Inc., <https://www.qiagenbioinformatics.com/products/ingenuitypathway-analysis>, (55)), focusing on endogenous regulators only. A list of 413 genes with preferential renal expression (53 kidney-enriched, 229 kidney-elevated, and 131 group-enriched) was extracted from Human Protein Atlas v19.3 ([www.proteinatlas.org](http://www.proteinatlas.org) (56)). Mendelian genes responsible for glomerulopathies (172) and tubulopathies (55) were extracted from Parsa et al. (57). A list of renal cell-type specific markers was curated from literature (58) as well as derived from single-cell RNA-seq data(59). Outcome association testing was performed using *survival* R package.

## References

1. B. Bikbov, C. A. Purcell, A. S. Levey, M. Smith, A. Abdoli, M. Abebe, O. M. Adebayo, M. Afarideh, S. K. Agarwal, M. Agudelo-Botero, E. Ahmadian, Z. Al-Aly, V. Alipour, A. Almasi-Hashiani, R. M. Al-Raddadi, N. Alvis-Guzman, S. Amini, T. Andrei, C. L. Andrei, Z. Andualem, M. Anjomshoa, J. Arabloo, A. F. Ashagre, D. Asmelash, Z. Ataro, M. M. d. W. Atout, M. A. Ayanore, A. Badawi, A. Bakhtiari, S. H. Ballew, A. Balouchi, M. Banach, S. Barquera, S. Basu, M. T. Bayih, N. Bedi, A. K. Bello, I. M. Bensenor, A. Bijani, A. Bloor, A. M. Borzì, L. A. Cámera, J. J. Carrero, F. Carvalho, F. Castro, F. Catalá-López, A. R. Chang, K. L. Chin, S.-C. Chung, M. Cirillo, E. Cousin, L. Dandonona, R. Dandonona, A. Daryani, R. Das Gupta, F. M. Demeke, G. T. Demoz, D. M. Desta, H. P. Do, B. B. Duncan, A. Eftekhari, A. Esteghamati, S. S. Fatima, J. C. Fernandes, E. Fernandes, F. Fischer, M. Freitas, M. M. Gad, G. G. Gebremeskel, B. M. Gebresillassie, B. Geta, M. Ghafourifard, A. Ghajar, N. Ghith, P. S. Gill, I. A. Ginawi, R. Gupta, N. Hafezi-Nejad, A. Haj-Mirzaian, A. Haj-Mirzaian, N. Hariyani, M. Hasan, M. Hasankhani, A. Hasanzadeh, H. Y. Hassen, S. I. Hay, B. Heidari, C. Herteliu, C. L. Hoang, M. Hosseini, M. Hostiuc, S. N. Irvani, S. M. S. Islam, N. Jafari Balalami, S. L. James, S. K. Jassal, V. Jha, J. B. Jonas, F. Joukar, J. J. Jozwiak, A. Kabir, A. Kahsay, A. Kasaeian, T. D. Kassa, H. G. Kassaye, Y. S. Khader, R. Khalilov, E. A. Khan, M. S. Khan, Y.-H. Khang, A. Kisa, C. P. Kovesdy, B. Kuate Defo, G. A. Kumar, A. O. Larsson, L.-L. Lim, A. D. Lopez, P. A. Lotufo, A. Majeed, R. Malekzadeh, W. März, A. Masaka, H. A. A. Meheretu, T. Miazgowski, A. Mirica, E. M. Mirrakhimov, P. Mithra, B. Moazen, D. K. Mohammad, R. Mohammadpourhodki, S. Mohammed, A. H. Mokdad, L. Morales, I. Moreno Velasquez, S. M. Mousavi, S. Mukhopadhyay, J. B. Nachega, G. N. Nadkarni, J. R. Nansseu, G. Natarajan, J. Nazari, B. Neal, R. I. Negoi, C. T. Nguyen, R. Nikbakhsh, J. J. Noubiap, C. Nowak, A. T. Olagunju, A. Ortiz, M. O. Owolabi, R. Palladino, M. Pathak, H. Poustchi, S. Prakash, N. Prasad, A. Rafiei, S. B. Raju, K. Ramezanzadeh, S. Rawaf, D. L. Rawaf, L. Rawal, R. C. Reiner, Jr., A. Rezapour, D. C. Ribeiro, L. Roeber, D. Rothenbacher, G. M. Rwegerera, S. Saadatagah, S. Safari, B. W. Sahle, H. Salem, J. Sanabria, I. S. Santos, A. Sarveazad, M. Sawhney, E. Schaeffner, M. I. Schmidt, A. E. Schutte, S. G. Sepanlou, M. A. Shaikh, Z. Sharafi, M. Sharif, A. Sharifi, D. A. S. Silva, J. A. Singh, N. P. Singh, M. M. M. Sisay, A. Soheili, I. Sutradhar, B. F. Teklehaimanot, B. e. Tesfay, G. F. Teshome, J. S. Thakur, M. Tonelli, K. B. Tran, B. X. Tran, C. Tran Ngoc, I. Ullah, P. R. Valdez, S. Varughese, T. Vos, L. G. Vu, Y. Waheed, A. Werdecker, H. F. Wolde, A. B. Wondmieneh, S. Wulf Hanson, T. Yamada, Y. Yeshaw, N. Yonemoto, H. Yusefzadeh, Z. Zaidi, L. Zaki, S. B. Zaman, N. Zamora, A. Zarghi, K. A. Zewdie, J. Ärnlöv, J. Coresh, N. Perico, G. Remuzzi, C. J. L. Murray, T. Vos, Global, regional, and national burden of chronic kidney disease, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017. *The Lancet* **395**, 709-733 (2020).
2. E. F. Carney, The impact of chronic kidney disease on global health. *Nature Reviews Nephrology* **16**, 251-251 (2020).
3. Chapter 1: Definition and classification of CKD. *Kidney Int Suppl (2011)* **3**, 19-62 (2013).
4. K.-U. Eckardt, J. Coresh, O. Devuyst, R. J. Johnson, A. Köttgen, A. S. Levey, A. Levin, Evolving importance of kidney disease: from subspecialty to global health burden. *The Lancet* **382**, 158-169 (2013).
5. C.-y. Hsu, G. M. Chertow, Chronic renal confusion: Insufficiency, failure, dysfunction, or disease. *Am. J. Kidney Dis.* **36**, 415-418 (2000).

6. C. Zoccali, R. Vanholder, Z. A. Massy, A. Ortiz, P. Sarafidis, F. W. Dekker, D. Fliser, D. Fouque, G. H. Heine, K. J. Jager, M. Kanbay, F. Mallamaci, G. Parati, P. Rossignol, A. Wiecek, G. London, R. on behalf of the European, A. Cardiovascular Medicine Working Group of the European Renal Association – European Dialysis Transplantation, The systemic nature of CKD. *Nature Reviews Nephrology* **13**, 344 (2017).
7. A. S. Levey, A. Levin, A Rebuttal to “The CKD Classification System in the Precision Medicine Era”. *Clin. J. Am. Soc. Nephrol.* **12**, 1711 (2017).
8. Y. N. Hall, J. Himmelfarb, The CKD Classification System in the Precision Medicine Era. *Clin. J. Am. Soc. Nephrol.* **12**, 346-348 (2017).
9. R. L. Luciano, G. W. Moeckel, Update on the Native Kidney Biopsy: Core Curriculum 2019. *Am. J. Kidney Dis.* **73**, 404-415 (2019).
10. S. K. Bhavnani, F. Eichinger, S. Martini, P. Saxman, H. V. Jagadish, M. Kretzler, Network analysis of genes regulated in renal diseases: implications for a molecular-based classification. *BMC Bioinformatics* **10 Suppl 9**, S3 (2009).
11. L. H. Mariani, W. F. Pendergraft, 3rd, M. Kretzler, Defining Glomerular Disease in Mechanistic Terms: Implementing an Integrative Biology Approach in Nephrology. *Clin. J. Am. Soc. Nephrol.* **11**, 2054-2060 (2016).
12. L. H. Mariani, M. Kretzler, Pro: 'The usefulness of biomarkers in glomerular diseases'. The problem: moving from syndrome to mechanism--individual patient variability in disease presentation, course and response to therapy. *Nephrol. Dial. Transplant.* **30**, 892-898 (2015).
13. C. Gluck, Y. A. Ko, K. Susztak, Precision Medicine Approaches to Diabetic Kidney Disease: Tissue as an Issue. *Curr. Diab. Rep.* **17**, 30 (2017).
14. T. Kohonen, in *Proceedings of International Conference on Neural Networks (ICNN'97)*. (1997), vol. 1, pp. PL1-PL6 vol.1.
15. P. Törönen, M. Kolehmainen, G. Wong, E. Castrén, Analysis of gene expression data using self-organizing maps. *FEBS Lett.* **451**, 142-146 (1999).
16. C. D. Cohen, M. Kretzler, [Gene expression analyses of kidney biopsies: the European renal cDNA bank--Kroner-Fresenius biopsy bank]. *Pathologe* **30**, 101-104 (2009).
17. J. P. Troost, J. Hawkins, D. R. Jenkins, D. S. Gipson, M. Kretzler, O. El Shamy, K. Bellovich, K. Perumal, Z. Bhat, S. Massengill, S. Steigerwalt, S. Pennathur, F. C. Brosius, C. A. Gadegbeku, Consent for Genetic Biobanking in a Diverse Multisite CKD Cohort. *Kidney Int Rep* **3**, 1267-1275 (2018).
18. C. A. Gadegbeku, D. S. Gipson, L. B. Holzman, A. O. Ojo, P. X. Song, L. Barisoni, M. G. Sampson, J. B. Kopp, K. V. Lemley, P. J. Nelson, C. C. Lienczewski, S. G. Adler, G. B. Appel, D. C. Cattran, M. J. Choi, G. Contreras, K. M. Dell, F. C. Fervenza, K. L. Gibson, L. A. Greenbaum, J. D. Hernandez, S. M. Hewitt, S. R. Hingorani, M. Hladunewich, M. C. Hogan, S. L. Hogan, F. J. Kaskel, J. C. Lieske, K. E. Meyers, P. H. Nachman, C. C. Nast, A. M. Neu, H. N. Reich, J. R. Sedor, C. B. Sethna, H. Trachtman, K. R. Tuttle, O.

- Zhdanova, G. E. Zilleruelo, M. Kretzler, Design of the Nephrotic Syndrome Study Network (NEPTUNE) to evaluate primary glomerular nephropathy by a multidisciplinary approach. *Kidney Int.* **83**, 749-756 (2013).
19. D. A. Muruve, M. C. Mann, K. Chapman, J. F. Wong, P. Ravani, S. A. Page, H. Benediktsson, The biobank for the molecular classification of kidney disease: research translation and precision medicine in nephrology. *BMC Nephrol.* **18**, 252 (2017).
  20. M. G. Sampson, J. B. Hodgins, M. Kretzler, Defining nephrotic syndrome from an integrative genomics perspective. *Pediatr. Nephrol.* **30**, 51-63; quiz 59 (2015).
  21. L. Braun, V. Sood, S. Hogue, B. Lieberman, C. Copley-Merriman, High burden and unmet patient needs in chronic kidney disease. *Int. J. Nephrol. Renovasc. Dis.* **5**, 151-163 (2012).
  22. C. M. Wyatt, D. Schlondorff, Precision medicine comes of age in nephrology: identification of novel biomarkers and therapeutic targets for chronic kidney disease. *Kidney Int.* **89**, 734-737 (2016).
  23. A. S. De Vriese, S. Sethi, J. Van Praet, K. A. Nath, F. C. Fervenza, Kidney Disease Caused by Dysregulation of the Complement Alternative Pathway: An Etiologic Approach. *J. Am. Soc. Nephrol.* **26**, 2917-2929 (2015).
  24. L. Sun, L. X. Zou, M. J. Chen, Make Precision Medicine Work for Chronic Kidney Disease. *Med. Princ. Pract.* **26**, 101-107 (2017).
  25. S. Lopez-Giacoman, M. Madero, Biomarkers in chronic kidney disease, from kidney function to kidney damage. *World J Nephrol* **4**, 57-73 (2015).
  26. M. Cañadas-Garre, K. Anderson, J. McGoldrick, A. P. Maxwell, A. J. McKnight, Genomic approaches in the search for molecular biomarkers in chronic kidney disease. *J. Transl. Med.* **16**, 292 (2018).
  27. M. Melé, P. G. Ferreira, F. Reverter, D. S. DeLuca, J. Monlong, M. Sammeth, T. R. Young, J. M. Goldmann, D. D. Pervouchine, T. J. Sullivan, R. Johnson, A. V. Segrè, S. Djebali, A. Niarchou, T. G. Consortium, F. A. Wright, T. Lappalainen, M. Calvo, G. Getz, E. T. Dermitzakis, K. G. Ardlie, R. Guigó, The human transcriptome across tissues and individuals. *Science* **348**, 660-665 (2015).
  28. M. Kretzler, W. Ju, A Transcriptional Map of the Renal Tubule: Linking Structure to Function. *J. Am. Soc. Nephrol.* **26**, 2603-2605 (2015).
  29. F. P. Schena, I. Nistor, C. Curci, Transcriptomics in kidney biopsy is an untapped resource for precision therapy in nephrology: a systematic review. *Nephrology Dialysis Transplantation* **33**, 1094-1102 (2017).
  30. M. Habuka, L. Fagerberg, B. M. Hallström, C. Kampf, K. Edlund, Å. Sivertsson, T. Yamamoto, F. Pontén, M. Uhlén, J. Odeberg, The Kidney Transcriptome and Proteome Defined by Transcriptomics and Antibody-Based Profiling. *PLoS One* **9**, e116125 (2015).

31. W. Ju, V. Nair, S. Smith, L. Zhu, K. Shedden, P. X. K. Song, L. H. Mariani, F. H. Eichinger, C. C. Berthier, A. Randolph, J. Y.-C. Lai, Y. Zhou, J. J. Hawkins, M. Bitzer, M. G. Sampson, M. Thier, C. Solier, G. C. Duran-Pacheco, G. Duchateau-Nguyen, L. Essioux, B. Schott, I. Formentini, M. C. Magnone, M. Bobadilla, C. D. Cohen, S. M. Bagnasco, L. Barisoni, J. Lv, H. Zhang, H.-Y. Wang, F. C. Brosius, C. A. Gadegbeku, M. Kretzler, C. P. N. the Ercb, P. K.-I. Consortium, Tissue transcriptome-driven identification of epidermal growth factor as a chronic kidney disease biomarker. *Sci. Transl. Med.* **7**, 316ra193-316ra193 (2015).
32. Y. Cui, R. S. Paules, Use of transcriptomics in understanding mechanisms of drug-induced toxicity. *Pharmacogenomics* **11**, 573-585 (2010).
33. S. Mihai, E. Codrici, I. D. Popescu, A.-M. Enciu, L. Albuлесcu, L. G. Necula, C. Mambet, G. Anton, C. Tanase, Inflammation-Related Mechanisms in Chronic Kidney Disease Prediction, Progression, and Outcome. *J Immunol Res* **2018**, 2180373-2180373 (2018).
34. B. B. Ratliff, W. Abdulmahdi, R. Pawar, M. S. Wolin, Oxidant Mechanisms in Renal Injury and Disease. *Antioxidants & Redox Signaling* **25**, 119-146 (2016).
35. D. Portilla, Apoptosis, Fibrosis and Senescence. *Nephron Clinical Practice* **127**, 65-69 (2014).
36. M. M. Speeckaert, R. Speeckaert, M. Laute, R. Vanholder, J. R. Delanghe, Tumor Necrosis Factor Receptors: Biology and Therapeutic Potential in Kidney Diseases. *Am. J. Nephrol.* **36**, 261-270 (2012).
37. A. A. Eddy, Progression in Chronic Kidney Disease. *Adv. Chronic Kidney Dis.* **12**, 353-365 (2005).
38. L. H. Mariani, S. Martini, L. Barisoni, P. A. Canetta, J. P. Troost, J. B. Hodgin, M. Palmer, A. Z. Rosenberg, K. V. Lemley, H. P. Chien, J. Zee, A. Smith, G. B. Appel, H. Trachtman, S. M. Hewitt, M. Kretzler, S. M. Bagnasco, Interstitial fibrosis scored on whole-slide digital imaging of kidney biopsies is a predictor of outcome in proteinuric glomerulopathies. *Nephrol. Dial. Transplant.* **33**, 310-318 (2018).
39. S. Kostadinova-Kunovska, G. Petrusevska, R. Jovanovic, L. Grcevska, M. Bogdanovska, M. Polenakovic, Morphological changes in the tubulointerstitial compartment in primary glomerulopathies. *Prilozi* **28**, 61-74 (2007).
40. A. Srivastava, R. Palsson, A. D. Kaze, M. E. Chen, P. Palacios, V. Sabbisetti, R. A. Betensky, T. I. Steinman, R. I. Thadhani, G. M. McMahon, I. E. Stillman, H. G. Rennke, S. S. Waikar, The Prognostic Value of Histopathologic Lesions in Native Kidney Biopsy Specimens: Results from the Boston Kidney Biopsy Cohort Study. *J. Am. Soc. Nephrol.* **29**, 2213-2224 (2018).
41. B. Rodríguez-Iturbe, R. J. Johnson, J. Herrera-Acosta, Tubulointerstitial damage and progression of renal failure. *Kidney Int. Suppl.*, S82-86 (2005).
42. S. Eddy, L. H. Mariani, M. Kretzler, Integrated multi-omics approaches to improve classification of chronic kidney disease. *Nat Rev Nephrol*, (2020).

43. S. Eddy, V. Nair, L. H. Mariani, F. H. Eichinger, J. Hartman, H. Huang, H. Parikh, J. N. Taroni, M. T. Lindenmeyer, W. Ju, C. S. Greene, P. C. Grayson, B. Godfrey, C. D. Cohen, M. G. Sampson, R. A. Lafayette, J. Krischer, P. A. Merkel, M. Kretzler, Inflammatory and JAK-STAT Pathways as Shared Molecular Targets for ANCA-Associated Vasculitis and Nephrotic Syndrome. *bioRxiv*, 427898 (2018).
44. L. H. Mariani, S. Eddy, S. Martini, F. Eichinger, B. Godfrey, V. Nair, S. G. Adler, G. B. Appel, A. Athavale, L. Barisoni, E. Brown, D. C. Cattran, K. M. Dell, V. Derebail, F. C. Fervenza, A. Fornoni, C. A. Gadegbeku, K. L. Gibson, D. Gipson, L. A. Greenbaum, S. R. Hingorani, M. A. Hlandunewich, J. Hogan, L. B. Holzman, J. A. Jefferson, F. J. Kaskel, J. B. Kopp, R. A. Lafayette, K. V. Lemley, J. C. Lieske, J.-J. Lin, K. E. Myers, P. H. Nachman, C. C. Nast, A. M. Neu, H. N. Reich, K. Sambandam, J. R. Sedor, C. B. Sethna, T. Srivastava, H. Trachtman, C. Tran, C.-s. Wang, M. Kretzler, Redefining Nephrotic Syndrome in Molecular Terms: Outcome-associated molecular clusters and patient stratification with noninvasive surrogate biomarkers. *bioRxiv*, 427880 (2018).
45. Y. Yasuda, C. D. Cohen, A. Henger, M. Kretzler, D. N. A. B. C. European Renal c, Gene expression profiling analysis in nephrology: towards molecular definition of renal disease. *Clin. Exp. Nephrol.* **10**, 91-98 (2006).
46. L. Barisoni, C. C. Nast, J. C. Jennette, J. B. Hodgins, A. M. Herzenberg, K. V. Lemley, C. M. Conway, J. B. Kopp, M. Kretzler, C. Lienczewski, C. Avila-Casado, S. Bagnasco, S. Sethi, J. Tomaszewski, A. H. Gasim, S. M. Hewitt, Digital pathology evaluation in the multicenter Nephrotic Syndrome Study Network (NEPTUNE). *Clin. J. Am. Soc. Nephrol.* **8**, 1449-1459 (2013).
47. R. Wehrens, L. M. C. Buydens, Self- and Super-organizing Maps in R: The kohonen Package. *Journal of Statistical Software; Vol 1, Issue 5 (2007)*, (2007).
48. L. Hamel, in *Advances in Self-Organizing Maps and Learning Vector Quantization*, E. Merényi, M. J. Mendenhall, P. O'Driscoll, Eds. (Springer International Publishing, Cham, 2016), pp. 49-59.
49. H. Löffler-Wirth, M. Kalcher, H. Binder, oposSOM: R-package for high-dimensional portraying of genome-wide expression landscapes on bioconductor. *Bioinformatics* **31**, 3225-3227 (2015).
50. J. Vesanto, E. Alhoniemi, Clustering of the self-organizing map. *IEEE Trans Neural Netw* **11**, 586-600 (2000).
51. M. Charrad, N. Ghazzali, V. Boiteau, A. Niknafs, NbClust: An R Package for Determining the Relevant Number of Clusters in a Data Set. *2014* **61**, 36 (2014).
52. M. E. Ritchie, B. Phipson, D. Wu, Y. Hu, C. W. Law, W. Shi, G. K. Smyth, limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47-e47 (2015).
53. A. Subramanian, P. Tamayo, V. K. Mootha, S. Mukherjee, B. L. Ebert, M. A. Gillette, A. Paulovich, S. L. Pomeroy, T. R. Golub, E. S. Lander, J. P. Mesirov, Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences* **102**, 15545 (2005).

54. V. K. Mootha, C. M. Lindgren, K.-F. Eriksson, A. Subramanian, S. Sihag, J. Lehar, P. Puigserver, E. Carlsson, M. Ridderstråle, E. Laurila, N. Houstis, M. J. Daly, N. Patterson, J. P. Mesirov, T. R. Golub, P. Tamayo, B. Spiegelman, E. S. Lander, J. N. Hirschhorn, D. Altshuler, L. C. Groop, PGC-1 $\alpha$ -responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat. Genet.* **34**, 267-273 (2003).
55. A. Kramer, J. Green, J. Pollard, Jr., S. Tugendreich, Causal analysis approaches in Ingenuity Pathway Analysis. *Bioinformatics* **30**, 523-530 (2014).
56. M. Uhlén, L. Fagerberg, B. M. Hallström, C. Lindskog, P. Oksvold, A. Mardinoglu, Å. Sivertsson, C. Kampf, E. Sjöstedt, A. Asplund, I. Olsson, K. Edlund, E. Lundberg, S. Navani, C. A.-K. Szigartyo, J. Odeberg, D. Djureinovic, J. O. Takanen, S. Hober, T. Alm, P.-H. Edqvist, H. Berling, H. Tegel, J. Mulder, J. Rockberg, P. Nilsson, J. M. Schwenk, M. Hamsten, K. von Feilitzen, M. Forsberg, L. Persson, F. Johansson, M. Zwahlen, G. von Heijne, J. Nielsen, F. Pontén, Tissue-based map of the human proteome. *Science* **347**, 1260419 (2015).
57. A. Parsa, C. Fuchsberger, A. Köttgen, C. M. O'Seaghdha, C. Pattaro, M. de Andrade, D. I. Chasman, A. Teumer, K. Endlich, M. Olden, M.-H. Chen, A. Tin, Y. J. Kim, D. Taliun, M. Li, M. Feitosa, M. Gorski, Q. Yang, C. Hundertmark, M. C. Foster, N. Glazer, A. Isaacs, M. Rao, A. V. Smith, J. R. O'Connell, M. Struchalin, T. Tanaka, G. Li, S.-J. Hwang, E. J. Atkinson, K. Lohman, M. C. Cornelis, Å. Johansson, A. Tönjes, A. Dehghan, V. Couraki, E. G. Holliday, R. Sorice, Z. Kutalik, T. Lehtimäki, T. Esko, H. Deshmukh, S. Ulivi, A. Y. Chu, F. Murgia, S. Trompet, M. Imboden, B. Kollerits, G. Pistis, T. B. Harris, L. J. Launer, T. Aspelund, G. Eiriksdottir, B. D. Mitchell, E. Boerwinkle, H. Schmidt, E. Hofer, F. Hu, A. Demirkan, B. A. Oostra, S. T. Turner, J. Ding, J. S. Andrews, B. I. Freedman, F. Giulianini, W. Koenig, T. Illig, A. Döring, H. E. Wichmann, L. Zgaga, T. Zemunik, M. Boban, C. Minelli, H. E. Wheeler, W. Igl, G. Zaboli, S. H. Wild, A. F. Wright, H. Campbell, D. Ellinghaus, U. Nöthlings, G. Jacobs, R. Biffar, F. Ernst, G. Homuth, H. K. Kroemer, M. Nauck, S. Stracke, U. Völker, H. Völzke, P. Kovacs, M. Stumvoll, R. Mägi, A. Hofman, A. G. Uitterlinden, F. Rivadeneira, Y. S. Aulchenko, O. Polasek, N. Hastie, V. Vitart, C. Helmer, J. J. Wang, B. Stengel, D. Ruggiero, S. Bergmann, M. Kähönen, J. Viikari, T. Nikopensius, M. Province, H. Colhoun, A. Doney, A. Robino, B. K. Krämer, L. Portas, I. Ford, B. M. Buckley, M. Adam, G.-A. Thun, B. Paulweber, M. Haun, C. Sala, P. Mitchell, M. Ciullo, P. Vollenweider, O. Raitakari, A. Metspalu, C. Palmer, P. Gasparini, M. Pirastu, J. W. Jukema, N. M. Probst-Hensch, F. Kronenberg, D. Toniolo, V. Gudnason, A. R. Shuldiner, J. Coresh, R. Schmidt, L. Ferrucci, C. M. van Duijn, I. Borecki, S. L. R. Kardia, Y. Liu, G. C. Curhan, I. Rudan, U. Gyllensten, J. F. Wilson, A. Franke, P. P. Pramstaller, R. Rettig, I. Prokopenko, J. Witteman, C. Hayward, P. M. Ridker, M. Bochud, I. M. Heid, D. S. Siscovick, C. S. Fox, W. L. Kao, C. A. Böger, Common Variants in Mendelian Kidney Disease Genes and Their Association with Renal Function. *J. Am. Soc. Nephrol.* **24**, 2105 (2013).
58. W. Ju, C. S. Greene, F. Eichinger, V. Nair, J. B. Hodgins, M. Bitzer, Y. S. Lee, Q. Zhu, M. Kehata, M. Li, S. Jiang, M. P. Rastaldi, C. D. Cohen, O. G. Troyanskaya, M. Kretzler, Defining cell-type specificity at the transcriptional level in human disease. *Genome Res.* **23**, 1862-1873 (2013).



59. H. Wu, A. F. Malone, E. L. Donnelly, Y. Kirita, K. Uchimura, S. M. Ramakrishnan, J. P. Gaut, B. D. Humphreys, Single-Cell Transcriptomics of a Human Kidney Allograft Biopsy Specimen Defines a Diverse Inflammatory Response. *J. Am. Soc. Nephrol.* **29**, 2069-2080 (2018).
60. M. Tomilo, H. Ascani, B. Mirel, M. C. Magnone, C. M. Quinn, A. Karihaloo, K. Duffin, U. D. Patel, M. Kretzler, Renal Pre-Competitive Consortium (RPC(2)): discovering therapeutic targets together. *Drug Discov Today* **23**, 1695-1699 (2018).

## **Acknowledgements**

We thank Dr. Lalita Subramanian for the excellent writing assistance and diligent proofreading of this manuscript. See Supplemental Acknowledgments for NEPTUNE consortium details.

ERCB members at the time of the study: Clemens David Cohen, Holger Schmid, Michael Fischereeder, Lutz Weber, Matthias Kretzler, Detlef Schlöndorff, Munich/Zurich/AnnArbor/New York; Jean Daniel. Sraer, Pierre Ronco, Paris; Maria Pia Rastaldi, Giuseppe D'Amico, Milano; Peter Doran, Hugh Brady, Dublin; Detlev Mönks, Christoph Wanner, Würzburg; Andrew Rees, Aberdeen and Vienna; Frank Strutz, Gerhard Anton Müller, Göttingen; Peter Mertens, Jürgen Floege, Aachen; Norbert Braun, Teut Rislér, Tübingen; Loreto Gesualdo, Francesco Paolo Schena, Bari; Gunter Wolf, Jena; Rainer Oberbauer, Donscho Kerjaschki, Vienna; Bernhard Banas, Bernhard Krämer, Regensburg; Moin Saleem, Bristol; Rudolf Wüthrich, Zurich; Walter Samtleben, Munich; Harm Peters, Hans-Hellmut Neumayer, Berlin; Mohamed Daha, Leiden; Katrin Ivens, Bernd Grabensee, Düsseldorf; Francisco Mampaso(†), Madrid; Jun Oh, Franz Schaefer, Martin Zeier, Hermann-Joseph Gröne, Heidelberg; Peter Gross, Dresden; Giancarlo Tonolo; Sassari; Vladimir Tesar, Prague; Harald Rupprecht, Bayreuth; Hermann Pavenstädt, Münster; Hans-Peter Marti, Bern; Peter Mertens, Magdeburg, Jens Gerth, Zwickau.

## **Funding**

This work was done as part of Renal Precompetitive Consortium (RPC2) collaboration (60) and was jointly funded by the participating members. This work is supported, in part, by George M. O'Brien Michigan Kidney Translational Core Center, funded by NIH/NIDDK grant P30-DK-081943. The Nephrotic Syndrome Study Network Consortium (NEPTUNE), U54-DK-083912, is a part of the National Institutes of Health (NIH) Rare Disease Clinical Research Network (RDCRN), supported through a collaboration between the Office of Rare Diseases Research, NCATS, and the National Institute of Diabetes, Digestive, and Kidney Diseases. Additional

funding and/or programmatic support for this project has also been provided by the University of Michigan, the NephCure Kidney International and the Halpin Foundation.

### ***Author contributions***

All co-authors have contributed to the manuscript. Specifically, A.R. conceived the project. A.R., M.C.M., M.K., V.N., R.B. designed and directed the project. A.R., V.N., S.E., M.T., T.S. analyzed the data and produced the figures. All authors interpreted the results, and provided critical feedback that helped shape the research, analysis and manuscript. A.R. drafted the manuscript. All authors approved the final version and accept accountability for the overall work by ensuring that questions pertaining to the accuracy or integrity of any portion of the work were appropriately investigated and resolved.

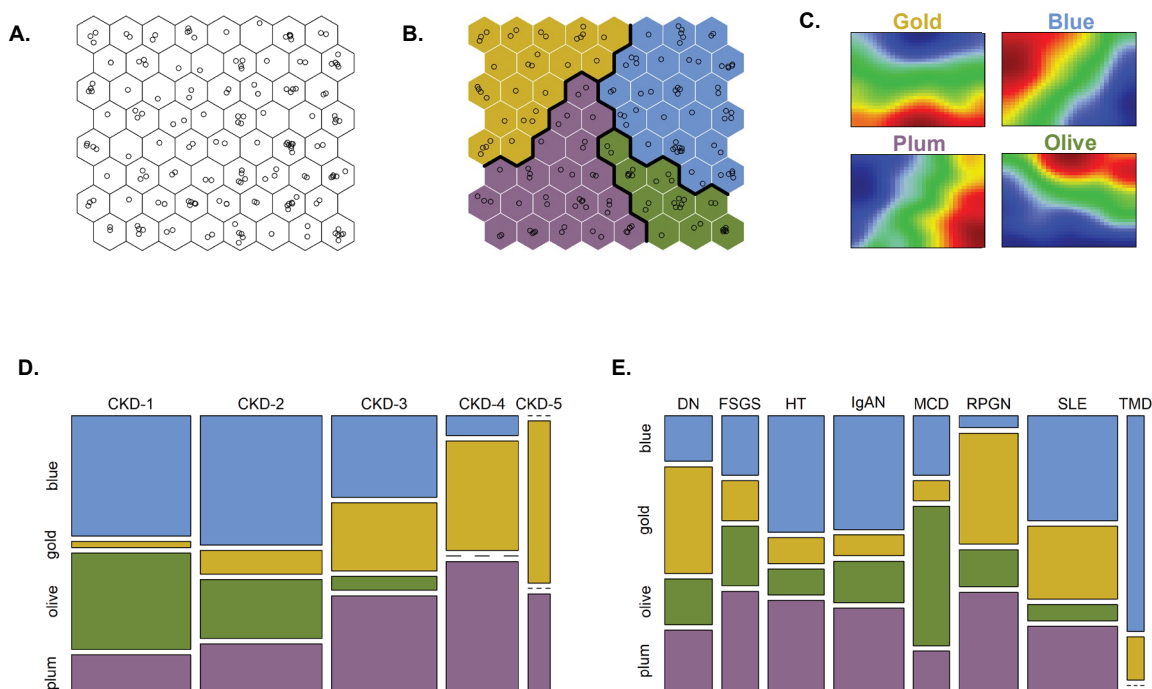
### ***Competing interests:***

Dr. Kretzler reports grants from NIH, non-financial support from University of Michigan, during the conduct of the study; grants from JDRF, Astra-Zeneca, NovoNordisk, Eli Lilly, Gilead, Goldfinch Bio, Merck, Chan Zuckerberg Initiative, Janssen, Boehringer-Ingelheim, Moderna, Chinook, amfAR, Angion, RenalytixAI, Retrophin, European Union Innovative Medicine Initiative and Certa outside the submitted work; In addition, Dr. Kretzler has a patent PCT/EP2014/073413 “Biomarkers and methods for progression prediction for chronic kidney disease” licensed. Dr. Eddy has served as co-Investigator on grants funded by Gilead Sciences, NovoNordisk, AstraZeneca, Janssen and Eli Lilly. Dr. Patel serves on the Board of Directors of Kidney Health Initiative. All other authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper. A.R., T.S., J.P.C., S.M., J.M.W., R.B. are employees of AstraZeneca. S.S.B., J.T.L., U.P.D. are employees and J.T.L is a shareholder of Gilead. C.M.Q. is an employee of Vifor Pharma, J.W., A.K. are employees of NovoNordisk. K.L.D. is an employee and stockholder of

Eli Lilly and Company. M.D.B., M.C.M. are employees, M.C.M is also a shareholder of Johnson and Johnson.

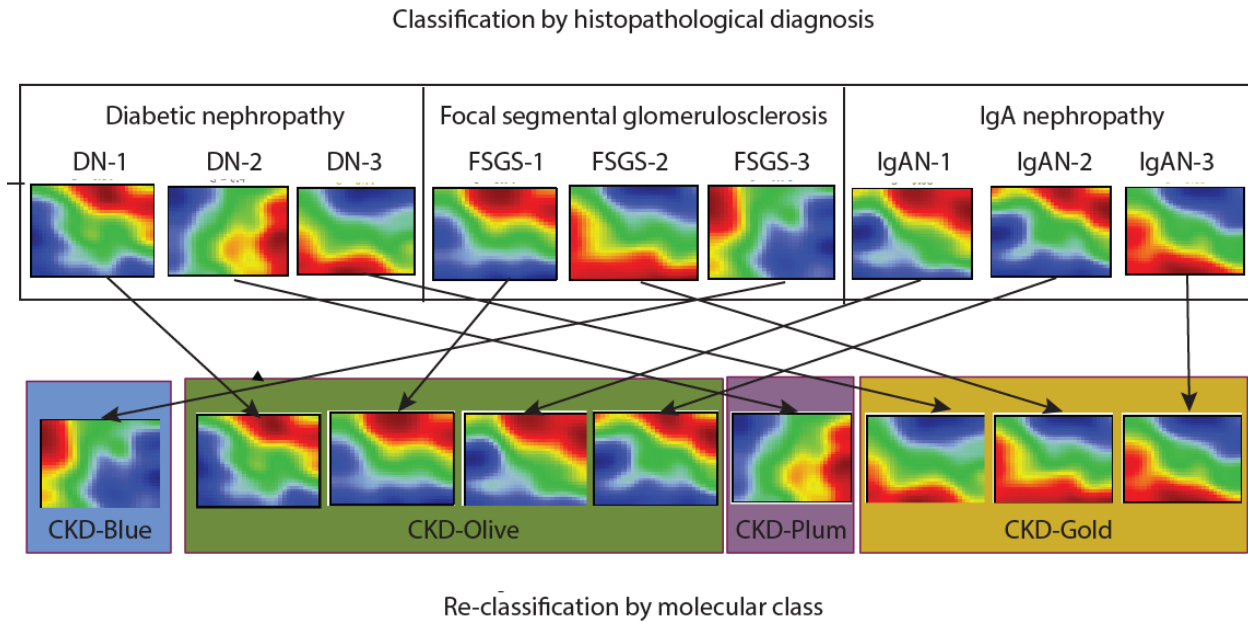
***Data and materials availability:***

All source data used in this study are publicly available and referenced in the Materials and Methods section in the main text. All data resulting from this study are available in the main text or the supplementary materials.



**Fig 1. Unbiased kidney transcriptomics stratification identified four inherent subgroups of patients (“molecular categories”) within a CKD cohort.**

Molecular categories were present at different CKD stages and histopathological diagnoses, providing orthogonal biomechanistic information regardless of the disease etiology or severity. **(A)** Self-Organizing Map of a CKD population based on kidney gene expression profiling. Individual patients (shown as open circles) were arranged in a topological order by similarities of their multivariable transcriptomics profiles. **(B)** Clustering of similar SOM units identified subgroups of similar patients. Thick lines indicate cluster boundaries. Clusters were assigned arbitrary colors. **(C)** Group-level summarized transcriptomics profiles (“expression portraits”) show distinct patterns between the molecular categories. Color scale reflects relative gene expression levels (red – high, blue – low, green – average) of 8,454 transcripts mapped. **(D-E)** Mosaic plots show the correspondence between molecular categorization and CKD stages and histopathological diagnoses, respectively. Width of the bands reflects the relative proportions of cases, the numbers of patients per category are also indicated by numbers.



**Fig 2. Heterogeneity of individual CKD transcriptomics profiles and re-classification based on molecular similarity independently from the conventional categories.**

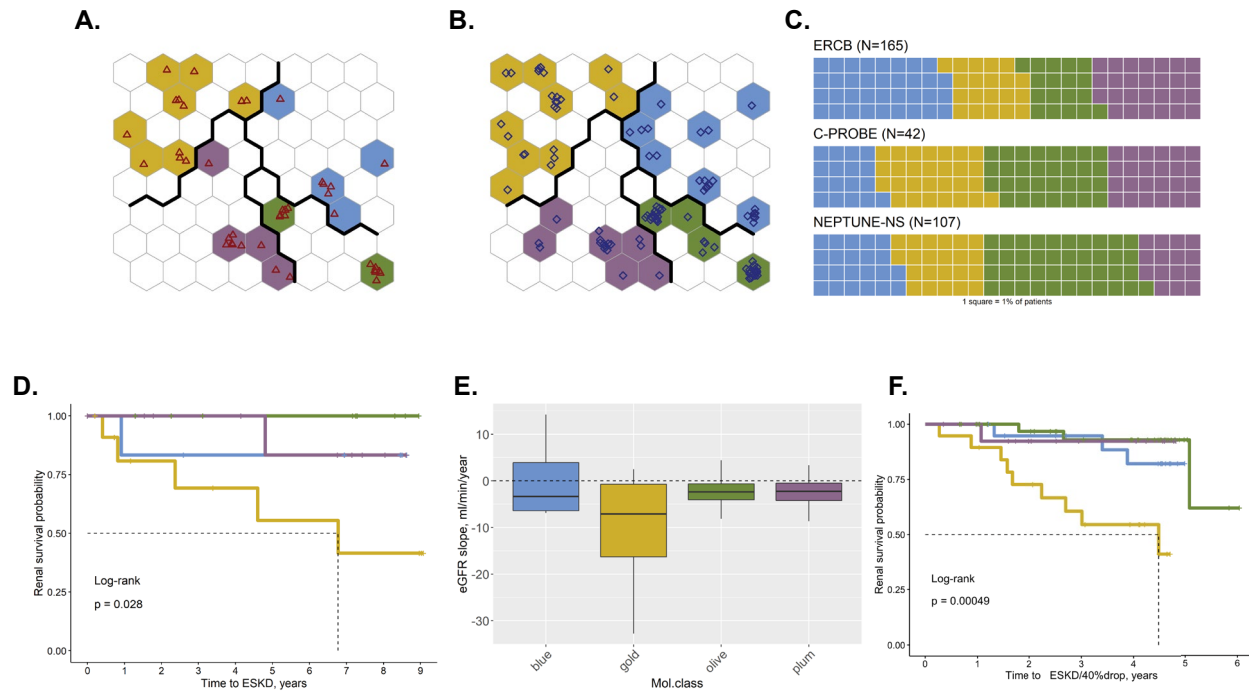
Sample of 9 patients' kidney gene expression profiles demonstrate marked heterogeneity, even within same diagnostic groups. Genes are represented in a fixed order on a grid to enable visual comparisons between samples. Color scale reflects gene expression levels relative to the population average (red – high, blue – low, green – average).



**Fig. 3. CKD molecular categories differ from healthy kidney transcriptomics profile and are biologically distinct.**

**(A)** Volcano plots visualize differential gene expression analysis results (log fold changes by statistical significance) per each molecular category contrasted with healthy control (living kidney donors, N=46). Each point represents a gene. Significantly modulated genes ( $q < 0.0001$ ) are highlighted in category-specific colors. Vertical dotted line indicates zero-fold change. The table shows numbers of differentially expressed genes, stratified by the direction of regulation, per each comparison. **(B)** Heatmap of gene set enrichment analysis results. The 5,205 gene sets (columns) were tested in each CKD molecular category (rows). The colors reflect normalized enrichment score (NES) values (maroon – positive, blue – negative) per gene set. The color bar indicates respective molecular category (Olive, Gold, Plum, Blue). Hierarchical clustering groups similar gene sets. **(C-D)** Venn diagrams visualize overlaps in significantly enriched (FDR  $q < 0.05$ ) up- and down-regulated gene sets, respectively, demonstrating presence of shared and unique pathways between CKD molecular categories. **(E)** Transcription regulator analysis. Activation Z-score for top 20 endogenous upstream regulators per CKD molecular category (orange – activated, green – inhibited). **(F)** Heatmap of hypothesis-driven enrichment analysis visualizing enrichments scores for CKD-relevant gene sets (kidney cell-type specific signatures, Mendelian genes known to cause renal phenotypes, genes with kidney-enriched expression). **(G)** Gene expression levels of known kidney injury markers in healthy controls and CKD molecular categories. Boxplots visualize median values per group, whiskers indicate IQR.





**Fig. 4. CKD molecular categorization in independent patient cohorts and disease progression comparison across CKD molecular categories.**

**(A)** C-PROBE cohort patients (N=42), mapped onto the trained discovery SOM; individual patients are shown as red triangles. **(B)** NEPTUNE-NS cohort patients (N=107) mapped onto the trained discovery SOM; individual patients are shown as blue diamonds. **(C)** Waffle charts visualize relative proportions of patients assigned to the different molecular categories in each CKD cohort (1 square=1%). **(D)** Kaplan-Meier curves of ESKD incidence by CKD molecular category in C-PROBE patients. Statistical significance of differences was tested using log-rank test. **(E)** Slope of eGFR, ml/min/year, by CKD molecular category in C-PROBE patients. Boxplots visualize median values per molecular class, whiskers indicate IQR. Dotted horizontal line denotes zero (no change in eGFR over time). **(F)** Kaplan-Meier curves of ESKD incidence by CKD molecular category in NEPTUNE patients. Statistical significance of differences was tested using log-rank test.

**Table 1.** Patient clinical characteristics for discovery and validation cohorts

Parameter	ERCB <sup>1</sup>	C-PROBE <sup>2,3</sup>	NEPTUNE <sup>2,3</sup>	Living donors <sup>4</sup>
N	165	42	107	46
Age, years	47±18	40±16	48±16	48±12
% Male	56%	38%	65%	50%
eGFR, ml/min/1.73m <sup>2</sup>	66±37	74±41	74±32	104±30
Proteinuria, g/24h	3.1 [0.7-4.2]	5.9 [5.1-6.5]	3.5 [1.6-6.95]	
<u>Histopathological diagnosis:</u>				
DN	17 (10%)	1 (2%)	.	
FSGS	13 (8%)	11 (26%)	46 (43%)	
HTN	20 (12%)	1 (2%)	.	
IgAN	25 (15%)	1 (2%)	.	
MGN	18 (11%)	2 (5%)	43 (40%)	
MCD	13 (8%)	1 (2%)	18 (17%)	
RPGN	21 (13%)	2 (5%)	.	
LN	32 (19%)	16 (38%)	.	
TMD	6 (4%)	.	.	
Other		7 (17%)	.	

<sup>1</sup>Discovery cohort; <sup>2</sup>Validation cohort; <sup>3</sup>Outcome association testing; <sup>4</sup>Control group for differential expression analyses

Quantitative variables are presented as means±SD or medians [IQR].

Abbreviations: eGFR – estimated glomerular filtration rate; DN - diabetic nephropathy; FSGS - focal segmental glomerulosclerosis; HTN - hypertensive nephrosclerosis; IgAN - immunoglobulin A nephropathy; MGN - membranous glomerulonephritis; MCD - minimal change disease; RPGN – rapidly progressive glomerulonephritis; LN - lupus nephritis; TMD – thin basement membrane disease.

**Table 2.** Summary of differential expression and enrichment gene sets by molecular CKD category

Parameter	CKD-Blue	CKD-Gold	CKD-Olive	CKD-Plum
Patients	58	33	29	45
Differentially expressed genes				
Up-regulated	45	1,756	386	1,660
Down-regulated	171	1,716	160	1,465
Significantly enriched gene sets				
Up-regulated	10	807	428	793
Down-regulated	122	343	483	140
Uniquely enriched gene sets				
Up-regulated	0	458	185	287
Down-regulated	56	272	376	27
Top 10 uniquely enriched up-regulated gene sets		<ul style="list-style-type: none"> <li>GO_adaptive_immune_response</li> <li>GO_extracellular_matrix_structural_constituent</li> <li>GO_leukocyte_proliferation</li> <li>GO_regulation_of_leukocyte_proliferation</li> <li>GO_positive_regulation_of_cell_activation</li> <li>GO_b_cell_receptor_signaling_pathway</li> <li>GO_membrane_invasion</li> <li>GO_humoral_immune_response</li> <li>Hallmark_inflammatory_response</li> <li>GO_positive_regulation_of_leukocyte_proliferation</li> </ul>	<ul style="list-style-type: none"> <li>GO_mitochondrial_gene_expression</li> <li>GO_mitochondrial_matrix</li> <li>GO_peroxisome_organization</li> <li>GO_mitochondrial_translational</li> <li>Reactome_peroxisomal_protein_import</li> <li>GO_peroxisomal_transport</li> <li>GO_cellular_amino_acid_metabolic_process</li> <li>GO_mitochondrial_protein_complex</li> <li>GO_ligase_activity</li> <li>GO_microbody</li> </ul>	<ul style="list-style-type: none"> <li>Reactome_mrna_splicing</li> <li>GO_nuclear_export</li> <li>Reactome_downstream_signaling_events_of_bcell_receptor_bcr</li> <li>GO_endosome_organization</li> <li>Reactome_scf_skp2_mediated_degradation_of_p27_p21</li> <li>Pid_nectin_pathway</li> <li>GO_gdp_binding</li> <li>Reactome_nod1_2_signaling_pathway</li> <li>GO_pigment_granule</li> <li>Kegg_spliceosome</li> </ul>
Top 10 uniquely enriched down-regulated gene sets	<ul style="list-style-type: none"> <li>GO_cellular_response_to_starvation</li> <li>GO_response_to_starvation</li> <li>Pid_foxo_pathway</li> <li>Pid_erbb1_downstream_pathway</li> <li>GO_kinase_inhibitor_activity</li> <li>Pid_tap63_pathway</li> <li>GO_dna_binding_transcription_factor_binding</li> <li>Hallmark_androgen_response</li> <li>GO_rna_polymerase_ii_specific_dna_binding_transcription_factor_binding</li> <li>GO_transcription_factor_binding</li> </ul>	<ul style="list-style-type: none"> <li>GO_mitochondrial_matrix</li> <li>GO_alpha_amino_acid_metabolic_process</li> <li>GO_coenzyme_binding</li> <li>Reactome_protein_localization</li> <li>GO_cofactor_binding</li> <li>GO_cellular_respiration</li> <li>GO_cellular_amino_acid_metabolic_process</li> <li>GO_mitochondrial_protein_complex</li> <li>GO_cellular_lipid_catabolic_process</li> <li>Reactome_mitochondrial_fatty_acid_beta_oxidation</li> </ul>	<ul style="list-style-type: none"> <li>GO_glomerulus_development</li> <li>Naba_secreted_factors</li> <li>Reactome_peptide_ligand_binding_receptors</li> <li>GO_neuropeptide_signaling_pathway</li> <li>Pid_fra_pathway</li> <li>GO_antimicrobial_humoral_immune_response_mediated_by_antimicrobial_peptide</li> <li>GO_calcium_channel_regulator_activity</li> <li>Reactome_formation_of_the_cornified_envelope</li> <li>GO_keratinization</li> <li>GO_mononuclear_cell_migration</li> </ul>	<ul style="list-style-type: none"> <li>GO_voltage_gated_potassium_channel_activity</li> <li>GO_glutamate_secretion</li> <li>GO_ligand_gated_ion_channel_activity</li> <li>GO_sodium_channel_activity</li> <li>GO_inorganic_ion_import_across_plasma_membrane</li> <li>Reactome_phase_0_rapid_depolarisation</li> <li>GO_transmission_of_nerve_impulse</li> <li>GO_potassium_ion_homeostasis</li> <li>Reactome_voltage_gated_potassium_channels</li> <li>GO_ligand_gated_cation_channel_activity</li> </ul>

## Supplementary Materials

**Supplemental Acknowledgement.** NEPTUNE Consortia members.

**Fig S1.** Gene expression levels across CKD molecular categories

**Fig S2.** eGFR by molecular category stratified by CKD stage

**Fig S3.** Kaplan-Meier survival curves for prediction of ESKD outcome stratified by CKD stage (A) and histopathological diagnosis (B)

**Fig S4.** Data clustering tendency assessment: pairwise Spearman correlation-based distances between samples (A); VAT, Visual Assessment of cluster Tendency (B)

**Fig S5.** Building Self-Organizing Map of CKD population: quantization error vs map size (A); convergence - embedding accuracy & topographic accuracy (B); learning progress across 10 independent models (C)

**Fig S6.** Final model SOM diagnostic plots: counts (A); mapping quality (B); neighbor distances (C); number of clusters (D)

**Table S1.** Differentially expressed genes per CKD molecular category vs healthy control

**Table S2.** Enriched pathways per CKD molecular category

**Table S3.** Upstream transcriptional regulators per CKD molecular category