Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

MEDA-PKD: A Dynamic, Real-Time Analytics Platform for Precision Care in Autosomal Dominant Polycystic Kidney Disease

Franziska Grundmann
University of Cologne
Sita Arjune
University of Cologne
Samer Alkarkoukly
University of Cologne
Malte Bartram
University of Cologne
Bodo Beck
University of Cologne
Thomas Benzing
University of Cologne
Volker Boehm
University of Cologne
Katrin Bohl
University of Cologne
Volker Burst
University of Cologne
Martin Busch
Jena University Hospital
Figen Cakiroglu
University Hospital Schleswig-Holstein
Peter Gerke
Nierenzentrum Lübeck
Nils Große-Hokamp
University of Cologne
Jan Halbritter
Charité - University Medicine Berlin
Ludwig Matrisch
University Hospital Schleswig-Holstein
Thorsten Persigehl

University of Cologne
Friederike Petzold
University Hospital Leipzig
Johannes Schmidt
Bonacci GmbH
Roland Schmitt
University Hospital Schleswig-Holstein
Polina Todorova
Heinrich Heine University Düsseldorf
Christoph Wanner
University Hospital Würzburg
Philipp Antczak
University of Cologne
Roman-Ulrich Müller
roman-ulrich.mueller@med.uni-duesseldorf.de

Heinrich Heine University Düsseldorf

Article

Keywords:

Posted Date: April 27th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9305978/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Additional Declarations: Competing interest reported. The Dept. 2 of Internal Medicine received research funding from Otsuka Pharmaceuticals, Alnylam, Vifor and ThermoFisherScientific and the AD(H)PKD Registry was co-funded by Otsuka Pharmaceuticals. RUM served as advisor to Alnylam, AICURIS, GSK, Vertex, Vifor and is a member of the Scientific Advisory Board of Santa Barbara Nutrients. NGH receives research funding and speakers fees from Philips Healthcare, speakers fees from Elsevier and serves as advisor to Bristol Myers Squibb and BeOne Medicines. RUM and JH served as chairs of the working group Genes&Kidney of the European Renal Association (ERA).

Version of Record: A version of this preprint was published at npj Digital Medicine on October 20th, 2026. See the published version at https://doi.org/10.1038/s41746-026-03267-0.

Abstract

Real-life medical data usage is currently limited by lack of structured databases and analysis tools. Autosomal dominant polycystic kidney disease (ADPKD) remains the most common monogenic cause of kidney failure. We present MEDA-PKD, an interactive, real-time, web-based analytics platform that integrates clinical routine with research data from a large, multicenter ADPKD cohort into a harmonized database (https://shiny.cecad.uni-koeln.de/ADPKD_registry/). Importantly, MEDA-PKD provides automated real-time analysis and dynamic visualization using ShinyApps, replacing traditional static approaches to clinical cohorts.

As of March 2026, MEDA-PKD visualized data from 1,735 patients (mean follow-up 796 days). Analysis of data from patients initiating tolvaptan therapy demonstrate the platform's capacity to evaluate therapeutic impact in real time. Other tabs include family history, medication, extrarenal manifestations, lifestyle, quality of life and a wide array of lab values (20,894 entries). Besides, MEDA-PKD integrates proteomics data with deep clinical phenotyping and allows for secure individual patient-level data access to participating centers.

MEDA-PKD represents a paradigm shift by transforming static cohort data into a living, actionable knowledge resource providing real-time insight in ADPKD and will serve as a template for other diseases.

Introduction

Chronic non-communicable diseases (NCD) are a major burden to society, patients, and their families. Chronic kidney disease (CKD) is a key example for these challenges. Autosomal dominant polycystic kidney disease (ADPKD) is the most common monogenic cause of CKD and kidney failure[1,2]. Formation of cysts begins in childhood, progresses over life-time and eventually replaces functional parenchyma leading to loss of kidney function. Kidney failure - which occurs by the sixth decade of life in about 50% of the patients - comes with a major burden both to ADPKD patients and healthcare systems[2,3]. Approximately 10% of all patients requiring kidney replacement therapy, i.e. kidney transplant or dialysis, have ADPKD as the underlying cause. Beside loss of kidney function, organ growth leads to various symptoms with high relevance to quality-of-life including pain and abdominal fullness[4]. As a systemic disorder, ADPKD also causes cyst formation in other organs, notably the liver with ~ 10–20% of the affected individuals presenting symptomatic polycystic liver disease (PLD)[5,6]. The disease also comes with an increased incidence of cerebral aneurysms (~ 8–12%), colonic diverticula, and cardiac valve defects[7].

More than 90% of all ADPKD cases are caused by pathogenic variants in two genes – *PKD1* and *PKD2*[8]. Despite its genetic nature, clinical outcomes in ADPKD including kidney failure show a high degree of intrafamilial variability[9] and cannot be accurately predicted yet. However, prediction of clinically relevant outcomes is of special importance in ADPKD to optimize patient counselling, implement individualized disease management and allow patients to determine the answers to major life questions. The only

treatment specifically approved for ADPKD, the vasopressin-V2-receptor antagonist tolvaptan[10], comes with relevant side effects including a risk of liver damage and severe polyuria (up to 15L / day)[11]. In addition, while tolvaptan slows loss of kidney function by about 25%[12,13], it will not prevent kidney failure in most cases. As to extrarenal manifestations, there are no pharmacological therapeutic options to date.

Consequently, a better characterization of the disease in a real-time fashion using large cohorts would be an important step to allow for (1) informing patients, physicians and other stakeholders about in-depth disease characteristics, (2) examining the impact of therapeutic interventions in a real-life setting, (3) leveraging the integration of clinical routine data with research data for the optimization of outcome prediction, (4) building the foundation for improved individual patient-counseling in the future and (5) allowing for direct patient stratification, e.g. as a basis for clinical trials. The usual approach to this topic is based on databases lacking standardization coupled to data cleaning and analysis in the context of individual projects. This approach results in static publication of visualized data allowing insight at a single-point in time and usually addressing a highly limited number of data points contained in the original list of variables collected. Furthermore, current standards usually require the manual transfer of data from clinical routine systems to research databases, a highly error-prone system further limiting the number of variables available to the analyses. However, thousands of variables are contained in clinical data systems and partly, e.g. for lab values, are available in structured databases already. MEDA-PKD addresses these challenges employing ADPKD as a highly relevant medical use-case. To reach this goal, MEDA-PKD integrates numerous data sources including direct access to laboratory values from the clinical database. Leveraging a PostgreSQL backend and R/Shiny-powered front-end visualization, MEDA-PKD provides clinicians and researchers with real-time insights into patient trajectories across key domains including eGFR dynamics, imaging phenotypes, extrarenal manifestations, medication use, quality of life, and family history. Built on a modular, containerized architecture across the shiny app as well as using Docker and Jenkins to perform underlying database updates, MEDA-PKD ensures long-term scalability, reproducibility, and interoperability. The system supports seamless integration of external datasets, enabling federated comparisons while preserving local data ownership.

The MEDA approach results in dynamic dashboards updating all analyses automatically after publication to provide the field with a real-time view into the characteristics of the cohort. Importantly, the dashboards already unify clinical data with omics-data in a first pilot approach. Users can filter for patient characteristics such as age, sex, tolvaptan usage and Mayo Imaging Class (MIC) within the individual dashboard to allow for an active user-webtool interaction for improved usability regarding individual research questions.

Taken together, MEDA-PKD is an important step forward for medical research in ADPKD. The data dictionary underlying MEDA-PKD will be provided to other centers to allow for extension to additional cohorts. Importantly, this data infrastructure is not limited to ADPKD but can essentially be employed for any disease setting after adaptation of the clinical variables required and definition of the research questions to design disease-specific dashboards.

Methods

AD(H)PKD cohort

Study Design and Cohort

This analysis is based on data from the German AD(H)PKD Tolvaptan Registry (NCT02497521, DRKS00008910), an ongoing multicenter, longitudinal observational study designed to collect real-world data from adults (≥ 18 years) with ADPKD. The registry systematically documents clinical, laboratory, imaging, and patient-reported data in individuals undergoing evaluation or treatment with tolvaptan in routine care or study follow-up.

All participants provided written informed consent prior to inclusion. The study protocol was approved by the Ethics Committee of the University of Cologne and all participating institutions. The study is conducted in accordance with the Declaration of Helsinki and the Good Clinical Practice (GCP) guidelines of the International Council for Harmonisation (ICH). The registry is publicly registered at ClinicalTrials.gov, NCT02497521 and the German Clinical Trials Register, DRKS00008910.

Inclusion Criteria

Eligible participants were adults (≥ 18 years) with clinically or genetically confirmed ADPKD who were followed at participating nephrology centers and evaluated for initiation or continuation of tolvaptan therapy.

Data Collection and Structure

Clinical data are collected prospectively and updated at least annually, with additional visits scheduled for patients initiating tolvaptan. Data capture follows a standardized electronic case report form using the secure web-based platform clinicalsurveys.net[14].

The dataset includes (but not exclusively):

- Demographics: age, sex, family history
- Genotype data
- Kidney function: serum creatinine, eGFR (CKD-EPI)
- Imaging parameters: total kidney volume (TKV) measured by MRI or CT and categorized by MIC.
- Medication: tolvaptan, antihypertensive medication, statins and opioids.
- Extrarenal manifestations: hepatic cysts, cardiovascular findings
- Lifestyle and patient-reported outcomes: smoking, alcohol consumption, and standardized quality-of-life questionnaires (SF-12)[15].

- Laboratory values: > 200 parameters covering electrolytes, liver function, metabolic, and inflammatory markers.

All data undergo automated quality control, plausibility checks, and harmonization across sites to ensure comparability and completeness.

Integration of Omics Data

To enable deep phenotyping and translational research, MEDA-PKD incorporates molecular data alongside clinical information. The current version includes a serum proteomics dataset from the Cologne AD(H)PKD subcohort[16]. Protein abundance data are continuously linked to clinical outcomes such as eGFR slope, MIC class, and treatment exposure. Users can select proteins of interest and visualize associations dynamically across patient subgroups (e.g., sex, CKD stage, tolvaptan use).

MEDA database and integration of diverse data sources

On the backend MEDA runs on a dockerized PostgreSQL database. Here all the data is stored, and backups generated daily stored securely off-site. To inject data into the database, several python scripts run at various intervals during the day. First, data is extracted for clinicalsurveys.net through their standalone application programming interfaces (APIs)[17]. These data are restructured and injected into the database. Next, routine clinical data is extracted from the medical information system of the hospital and imported. Finally, to enable additional user entry in a familiar style to clinicians we employ a NocoDB instance linked to the PostgreSQL database. This provides a web-based table interface and enables data entry of information that is not captured by the remaining systems. In addition, the shiny interface, introduced in the next section, also has the ability to import more static data, such as OMICs data or other associated datasets.

MEDA visualization

The MEDA-PKD platform was designed to provide an interactive, real-time environment for the visualization and analysis of longitudinal ADPKD registry data. Visualization modules were developed using R/Shiny, integrating the ggplot2 and plotly libraries to enable dynamic generation of time-series plots, stratified comparisons, and correlation analyses across multiple clinical domains. Data from the PostgreSQL backend are accessed through standardized application programming interfaces (APIs), allowing real-time rendering of graphics and ensuring consistency and reproducibility of all outputs.

The platform enables stratified visualization of clinical trajectories according to sex, age group, CKD stage, MIC, tolvaptan treatment status, or genotype. Within the kidney function dashboard, longitudinal eGFR curves and estimated slopes are visualized for individual patients as well as aggregated subgroups, providing an intuitive overview of disease progression. Imaging parameters, including total kidney volume (TKV), can be examined by MIC class and related directly to renal function decline.

Medication use, adherence, and discontinuation are visualized over time to facilitate assessment of treatment effects in real-world settings. Extrarenal manifestations such as hepatic cysts or cardiovascular anomalies are represented as frequency distributions, while patient-reported outcomes—including lifestyle factors and standardized quality-of-life measures—are displayed in longitudinal and cross-sectional formats.

A dedicated laboratory explorer allows for interactive visualization of 214 routinely measured parameters through customizable comparative boxplots, supporting the identification of biochemical patterns across treatment and disease subgroups. Advanced analytics functions enable users to perform subgroup analyses and regression modeling directly. For example, the eGFR Slope Explorer calculates and visualizes the annual decline in kidney function for selected patient groups, with options to overlay treatment status or biomarker levels. The omics integration module connects molecular data, such as proteomics, with clinical trajectories and automatically updates associations as new data are added to the registry.

All visualization outputs can be exported as static or interactive figures containing metadata to ensure transparency and traceability. Analytical scripts used within the platform are version-controlled via GitHub and automatically deployed through Jenkins within a continuous integration framework, guaranteeing reproducibility of all visual and analytical processes. By transforming static cohort data into an interactive analytical ecosystem, MEDA-PKD empowers clinicians and researchers to generate hypotheses, evaluate therapeutic responses, and explore biomarker-disease relationships in real time, thereby advancing precision medicine in ADPKD.

Statistical analyses

Statistical analyses in MEDA are governed by the underlying data distribution. Group comparisons used chi-square or Fisher's exact tests for categorical variables (e.g., sex differences in tolvaptan use) and one-way ANOVA or Kruskal-Wallis for continuous variables (e.g., eGFR slopes by MIC class), with post-hoc Tukey or Dunn's tests as appropriate. Two-group continuous comparisons used unpaired t-tests (normal) or Mann-Whitney U (non-normal). Values that are not normally distributed are log transformed prior to their analysis. For the predictive model, a linear regression model for eGFR slope from Aydogan-Balaban et. al.[16] is integrated directly into MEDA. eGFR slopes are calculated by a simple robust linear model predicting eGFR as a function of date. At least three values over a one-year period are required for a slope to be calculated. The resulting coefficient is then multiplied by 365.25 to represent an annual eGFR slope for an individual. Tolvaptan slopes are compared through a simple t-test and p-values reported. Fitted lines generated within the laboratory values over time, are based on a loess smoothed curve with standard parameters. Missing data are removed prior to analysis and all sample sizes are reported.

Data Sharing Statement

The full proteomics and clinical data supporting the findings of this study are not publicly archived to ensure data protection of study participants and minimize the risk of re-identification. Access to the dataset may be granted upon direct request to the corresponding authors depending on the nature of research questions aligning with the aims of the study to which participants provided consent and the ability to ensure data protection. To ensure this goal, data sharing will depend on a signed bilateral data transfer agreement. Interested researchers should contact the corresponding authors at roman-ulrich.mueller@uk-koeln.de or philipp.antczak@uk-koeln.de, including a description of the intended use and any institutional affiliations. Alternatively, analyses performed by the coordinating centers and provision of aggregated can be requested using the same email contacts.

Results

Baseline characteristics of the AD(H)PKD cohort

The technical backbone of the MEDA platform has recently been published by our group[17] and is now applied to ADPKD as a clinically relevant use-case with a defined need for real-life data analysis. The specific set of data sources fed into MEDA for this use-case includes research databases, the clinical information system ORBIS and omics data (for an overview see Suppl. Figure 1). The figures provided in the main body of this manuscript depict the static view on the visualizations as of March 3rd, 2026. MEDA itself is accessible at https://shiny.cecad.uni-koeln.de/ADPKD_registry/ and provides data visualizations, which are updated automatically on a daily basis. The graphs are provided in 13 dashboards dedicated to individual topics of interest (Suppl. Figure 2A). Data entering the graphs can be filtered by the user to obtain an insight into clinically relevant subgroups (e.g. age, sex, tolvaptan usage, MIC, Suppl. Figure 2B). The AD(H)PKD cohort study was set up in 2015 to further characterize natural disease course and the impact of the newly approved targeted therapeutic tolvaptan. AD(H)PKD recruits patients with ADPKD, confirmed by genetics or imaging criteria, with an eGFR ≥ 15 ml/min/1.73m^2. As of March 3rd, 2026, the study had recruited 1750 patients with sufficient data availability for MEDA-PKD in 1735 patients, 52% of whom were female and 48% male (Table 1, Fig. 1A, Suppl. Figure 3, Suppl. Figure 4A). 80.4% of patients reported a positive family history (Fig. 1A, Suppl. Figure 5). Reported median age was 44 years at baseline (Table 1, Suppl. Figure 3). 7% were already on tolvaptan at baseline and 26% reported being on or having taken tolvaptan at any of the visits (Table 1, Fig. 1A, Suppl. Figure 3). Median age at diagnosis was 27 years and mean GFR was 69 ml/min/1.73m^2 (Table 1, Fig. 1B, Suppl. Figure 3, Suppl. Figure 4B). 12% and 4% reported a very early onset of dialysis or kidney transplantation before the age of 50, respectively, in the affected parent (Suppl. Figure 5).

Table 1
Baseline characteristics of the AD(H)PKD cohort (March 3rd, 2026: N = 1735)

Age (years)	**Mean**	**SD**	**Median**	**Min**	**Max**	**Missing**
	43.8	**12.8**	**44.0**	**18.0**	**83.0**	
Age at Diagnosis (years)	28.0	13.8	27.0	0.0	83.0	73 (4.2%)
Weight (kg)	82.7	19.9	80.0	40.0	189.0	26 (1.5%)
Height (cm)	175.0	11.8	175.0	61.0	209.0	23 (1.3%)
Systolic BP (mmHg)	140.0	18.3	138.0	18.0	228.0	56 (3.2%)
Diastolic BP (mmHg)	87.9	12.8	87.0	33.0	144.0	56 (3.2%)
eGFR (ml/min/1.73m³)	69.3	31.9	67.2	5.6	143.0	24 (1.4%)
htTKV (ml/m)	1010	843	767	132	9870	176 (10.1%)
	Female (%)		**Male (%)**			
Sex	895 (51.6%)		840 (48.4%)			

Extrarenal manifestations

Patients were also asked about or examined for key extrarenal manifestations. It is important to note, that, in a real-life study setting and in line with the KDIGO guideline, no routine screening e.g. for intracranial aneurysms was performed. The vast majority of patients showed liver cysts (82% in female and 70% in male patients, Fig. 1D/E). Arachnoid (1% in female and 2% in male patients), splenic (2% in female and 1% in male patients), pancreatic (8% in female and 6% in male patients) and seminal vesicle cysts (2% in male patients) were reported at a much lower prevalence (Fig. 1D/E). A large proportion of patients showed cardiac valve abnormalities, the vast majority of which were mild in nature. Mitral regurgitation was the most common finding affecting 21% in female and 19% in male patients (Suppl. Figure 4C). Intracranial aneurysms had been detected in ~ 2% of male and female patients and hemorrhage had occurred in ~ 1%, again without differences between male and female participants (Suppl. Figure 4D).

General medication usage and lifestyle

78% of participants were on at least one antihypertensive drug. ACE-inhibitors or AT1 receptor antagonists were the most commonly used drug classes for this purpose (Suppl. Figure 6A). 2% of patients reported regular intake of non-opioid pain medication and 1% reported opioid usage for pain control (Suppl. Figure 6B). As to lipid lowering drugs, 14% of participants were on a statin as long-term medication (Suppl. Figure 6C). 19% of patients reported to be active-smokers with ~ one third reporting < 20 pack years, 8% 20–40 pack years and 2% ≥ 40 pack years (Suppl. Figure 7A/B). 12% of participants did not drink any coffee. 56% drank an equivalent of 1–3 cups per day and 33% reported a coffee

consumption of > 3 cups a day (Suppl. Figure 7C). Alcohol consumption was reported by 71% with only 4% consuming ≥ 25 g of alcohol per day (Suppl. Figure 7D).

Predictors of disease progression

As expected, total kidney volume increased with patient age and male patients showed larger kidneys than female patients (Suppl. Figure 8A). MIC was available for > 90% of participants (Fig. 2A). 61% of patients for whom a volumetry was available (n = 1574) showed a MIC indicating rapid progression (1C-1E) and 4% had been classified as showing atypical cyst distribution / morphology, i.e. MIC 2 (Fig. 2A). 41.5% of those with an available genotype (n = 562) showed a truncating PKD1 variant, 10.3% a non-truncating PKD1 variant and 18.1% a variant in the PKD2 gene (counting ACMG class 4 or 5 variants only; Fig. 2B), while the remainder contained minor PKD genes as well as variants of uncertain significance and biallelic cases. 65% of the entire cohort had experienced urological complications as defined in the PROPKD score (Suppl. Figure 8B) and among patients ≥ 35 years of age these had first manifested before the age of 35 in 16% (Suppl. Figure 8D). Flank pain was the most common urological complication followed by macrohematuria (Suppl. Figure 8F). Arterial hypertension was known in 84% of participants (Suppl. Figure 8C) and, in patients > 35 years of age, had first been diagnosed before the age of 35 in 24% (Fig. 8F). Based on these data, the PROPKD score could be calculated for 298 patients (i.e. genotype available and age ≥ 35 years) and indicated intermediate or high risk of rapid progression in 47 and 16% of participants, respectively, with male patients being more likely to be in the high-risk group (Fig. 2C).

Kidney function trajectories

As of March 3rd, 2026, 18,841 creatinine values were available for the AD(H)PKD cohort. Since tolvaptan as the only available disease-modifying treatment is known to have a clear impact on eGFR slopes only values which were obtained before starting tolvaptan were used to obtain an insight into natural disease progression characteristics. Using these values was sufficient to calculate eGFR slopes for 992 patients in a therapy-naïve state. Of these 958 passed the criteria of being below 5 and above – 15 ml/min/year. These analyses showed a faster eGFR decline in male patients (-3.2 ml/min/year compared to -2.8 ml/min/year in female patients, Fig. 2D). Annual kidney function loss increased with increasing MIC as expected (1A -2.0 ml/min/year, 1B -2.2 ml/min/year, 1C -3.0 ml/min/year, 1D -3.7 ml/min/year, 1E -4.3 ml/min/year, Fig. 2E). MIC 2 patients showed a mean eGFR slope compatible with slow disease progression (-1.8 ml/min/year; Fig. 2E). Besides showing aggregated data, MEDA can be used to provide study centers with individual patient-level data after secure dual authentication login (Suppl. Figure 9A). This is especially useful for visualizing individual eGFR slopes, e.g. to assess timing towards future onset of kidney failure, in a way ensuring data security (Suppl. Figure 9B and C, also see section on “individual patient dashboard below).

Usage of and adherence to tolvaptan in the real-life setting

The majority of patients were not on tolvaptan and did not start tolvaptan during follow-up. ~26% of patients took tolvaptan at any point during data collection and - out of these – 9% had stopped tolvaptan during follow-up (Fig. 3B). In these 166 patients, polyuria / nocturia was the most common reason for discontinuation (27%) followed by reaching an eGFR of < 15 ml/min/1.73m^2 (27%; Fig. 3C). Transaminase elevation was reported as the cause for stopping tolvaptan in 9% of all patients that discontinued the treatment (Fig. 3C).

Effect of tolvaptan on kidney function decline and quality of life in the real-life setting

We performed a longitudinal paired analysis of the impact of tolvaptan on eGFR decline in all patients for whom sufficient creatinine data was available over a period of at least one year before taking tolvaptan and while being on treatment. This approach showed a highly significant amelioration of kidney function loss by approximately 27% (p-value 0.00015; Fig. 3A). When assessing the slope differences in a sex specific manner both reached statistical significance, with men showing greater response than women (men: p-value 0.0012, women: p-value 0.035 Suppl. Figure 10A/B). Quality of life was assessed using the SF-12 tool. Interestingly, while the physical health score was comparable, the mental health score appeared to be better on tolvaptan when comparing all questionnaires filled on tolvaptan compared to those filled off tolvaptan (Suppl. Figure 11A/B). The same holds true when comparing the first questionnaire completed by patients who never took tolvaptan to the first questionnaire of patients filled after starting tolvaptan (Suppl. Figure 1C/D). However, longitudinal paired analysis for all patients that had filled at least one questionnaire before and on tolvaptan showed no tolvaptan-dependent differences in both the mental and the physical health score (Fig. 3D/E).

Individual patient dashboard enabling a patient-specific overview of ADPKD-related key characteristics

To allow for insight into the effect of disease progression and the impact of tolvaptan in individual patients, we established an additional dashboard. This dashboard is not freely available to respect data protection on a single-patient level. Access is provided to participating centers through a dual authentication-based login (Suppl. Figure 9A). Once logged in, treating clinician-scientists can select individual patients from their cohort to get a full overview of key characteristics important to patient in care in ADPKD summarized on one page. This includes an overview of all visits at the center visualizing phases with tolvaptan treatment and respective dosage (Fig. 4H). In addition, key characteristics are shown including sex, age, age at diagnosis, blood pressure and CKD stage (Fig. 4B). Extrarenal cysts are visualized on a body map (Fig. 4A) and parameters predicting outcome are summarized including MIC, BMI, PROPKD score, genotype and eGFR slope (based on values without tolvaptan, Fig. 4C). Furthermore, the dashboard includes family history (Fig. 4F) and urological symptoms (Fig. 4G). Importantly, eGFR slope is visualized and shown separately for phases with and without tolvaptan allowing to assess response and to extrapolate the time until kidney failure (Fig. 4E and 4I). The user can select individual lab values from the full set (Suppl. Figure 12) which are then visualized in the dashboard over time

(Fig. 4J). Key concomitant medication is shown by classes (Fig. 4D) providing a quick insight into whether relevant treatments have been implemented.

Integrating omics-data with clinical data to identify biomarkers for relevant outcome parameters

We have recently published a serum proteomics dataset on a substantial proportion of the AD(H)PKD cohort identifying a set of proteins predicting eGFR slope[16]. MEDA-PKD now allows the user to query this dataset to examine the association of proteins of choice with markers of disease severity in ADPKD linking omics-data to clinical parameters in a user-defined fashion. The user can define any protein identified in the dataset and visualize its abundance across different levels of eGFR or different MIC (shown for GPX3 in an exemplary manner in Fig. 5A). Besides, any protein of choice can be added to a heatmap in which clinical parameters can be added freely by the user (Fig. 5B). Most importantly, while the original publication provides the static model performance of a 6-protein model (including SERPINF1, GPX3, AFM, FERMT3, CFHR1 and RARRES2) MEDA-PKD now links this model to the actual real-time slopes from the cohort and visualizes calculated vs. predicted eGFR slope (Fig. 5C). To not limit biomarker identification to research omics-data but also allow the user to explore the relation of classical clinical laboratory values to markers of disease severity, MEDA-PKD features a “Lab Value” dashboard allowing the user to access 214 laboratory parameters. Graphs showing the impact of tolvaptan on urine volume and urine osmolarity as well as transaminases are always shown in the dashboard as standard parameters (Suppl. Figure 12A). Besides, the user can select any of the 214 variables to be visualized by sex, MIC, CKD stage and tolvaptan status.

Discussion

The MEDA-PKD platform makes an important step toward real-time precision nephrology and illustrates how routinely collected clinical and molecular data can be transformed into a continuously learning ecosystem. By linking longitudinal registry data with hospital information systems, imaging, and omics resources through an automated, harmonized infrastructure, MEDA-PKD transcends the limitations of traditional static registries. Rather than providing single snapshots of aggregated data, it enables dynamic exploration of disease trajectories and treatment effects as data accrue, closing the long-standing gap between data generation, analysis, and clinical translation[18–20].

From a clinical perspective, the results generated within MEDA-PKD confirm established patterns of ADPKD progression, such as sex- and TKV-based differences in eGFR decline[21,22], while providing the granularity to detect therapy effects under real-world conditions. The observed 27% reduction in eGFR loss on therapy with tolvaptan - obtained through an automated longitudinal update rather than retrospective reanalysis – is in line with the effect size observed in the pivotal randomized clinical trials (RCTs)[12,13] and underscores the platform’s capacity to evaluate therapeutic interventions in near real time. Importantly, integration of patient-reported outcomes and lifestyle variables extends the

interpretability of these findings beyond renal endpoints, offering a more comprehensive view of disease burden and treatment benefit. This capacity to continuously evaluate both biological and patient-centered outcomes is essential to accelerate precision medicine in chronic kidney disease and other complex disorders[23,24]. In this regard, the integration of total liver volume (TLV)-associated progression groups in patients affected by polycystic liver disease would be of interest[25]. Furthermore, different layers of outcome prediction in ADPKD could be integrated as recently performed for PROPKD score and MIC[26].

Technically, MEDA introduces several key innovations that differentiate it from existing registry or data warehouse approaches. The platform's modular architecture, combining a dockerized PostgreSQL backend with an R/Shiny-based analytics interface, ensures scalability, reproducibility, and interoperability. Automated data ingestion from multiple sources - including clinical information systems, research databases, and omics data in a unique manner - minimizes manual curation while maximizing data availability and reduces error-prone data transfer. Continuous integration via Jenkins and version control through GitHub guarantee that every analytical and visualization process is transparent, traceable, and reproducible. These design principles enable MEDA to function as a "living" data resource in which new information directly enriches existing analyses without additional manual intervention[17].

Compared to established infrastructures such as the ERA and ERKReg kidney disease registries[27,28], or large-scale frameworks like CKDGen[29] and the UK Biobank[30], MEDA differs in both granularity and immediacy. Most existing registries operate on static, periodic data exports and require centralized re-analysis to generate new insights. While these efforts have provided invaluable large-scale population data, their architectures are not designed for real-time interaction or automated integration of multi-omic datasets. In contrast, MEDA enables clinicians and researchers to dynamically explore longitudinal data, stratify patient subgroups, and link molecular signals to clinical phenotypes without additional programming effort. This immediacy creates an unprecedented opportunity for rapid hypothesis generation, validation of clinical observations, pre-screening for clinical trials and identification of therapeutic response patterns under routine care conditions.

This framework will be of benefit to many stakeholders in the field and can serve as a basis for hypothesis generation and validation to scientists, for trial design to pharmaceutical industry and for patient information and involvement through lay language versions. In this regard, the inclusion of patient-reported outcome measures (PROMs) such as the SF12 questionnaire is of special interest. The dual authentication protected internal version allows individual centers to visualize single-patient level data, e.g. to inform treatment decisions or as a useful addition patient counseling. Real-time visualization of treatment responses makes real-life cohort data accessible for an assessment of treatment efficaciousness, e.g. valuable to policymaking in the healthcare reimbursement sector, and the impact of previously unstudied therapeutic combinations or subgroup analyses. Moreover, this approach provides insight in to real-life tolerability of side effects, adherence and causes of treatment cessation[11].

Beyond the specific context of ADPKD, the MEDA framework provides a scalable blueprint for translational data infrastructures in other diseases. Its data model and dashboard modules are disease-agnostic and can be adapted to different clinical contexts by modifying variable dictionaries and analytics templates. For example, the same technical backbone could support cohorts focusing e.g. on diabetes mellitus associated CKD, but also heart failure, cystic fibrosis, or cancer. Extension to existing cohort platforms such as ERKReg and integration with data infrastructures such as the German National Research Data Infrastructure (NFDI4Health)[31] or the German Portal for Medical Research Data (FDPG)[32] is readily achievable via standardized APIs and data harmonization pipelines. By supporting federated analyses while preserving local data ownership, MEDA addresses key regulatory and ethical barriers that have historically limited multicenter data integration[33,34].

In the broader landscape of digital health research, MEDA contributes a unique combination of interoperability, automation, and interpretability. Whereas most hospital data warehouses or FAIR-compliant research databases focus on data availability[35], MEDA emphasizes accessibility and usability turning data into directly interpretable, hypothesis-generating visualizations. The open-source analytical scripts and modular container design also ensure sustainability: the platform can be maintained, extended, and reused across institutions without dependence on proprietary software ecosystems. In this way, MEDA bridges the traditional divide between clinical registries, translational research platforms, and clinical decision support tools[36,37] and makes an important contribution to FAIR data sharing strategies. Besides, external centers can submit their data after variable mapping in a harmonized format to build cohort-specific subtools and provide access to individual patient dashboard views.

Several limitations and future perspectives warrant consideration. First, while the platform provides continuous automated data ingestion and quality control, completeness and accuracy remain dependent on underlying source systems. Ongoing harmonization efforts and increasing adoption of structured electronic health records will further enhance data fidelity. Second, current implementations of the omics integration module focus on proteomics[16]; expansion to genomics, metabolomics, and imaging-derived radiomics is underway[38]. Finally, sustained engagement by clinicians and researchers will be essential to realize the full potential of MEDA as a community resource.

In conclusion, MEDA-PKD exemplifies how digital infrastructure can transform medical cohorts into dynamic, self-updating ecosystems for precision medicine. Its application spans clinical care, trial enrichment, biomarker discovery, and post-marketing drug surveillance. MEDA-PKD provides a scalable, interoperable, and transparent platform that not only advances the understanding and management of ADPKD but also offers a generalizable template for other diseases. By coupling real-world clinical data with automated analytics and translational integration, MEDA moves beyond static registry concepts toward a continuously learning health system -a cornerstone for next-generation biomedical research[39].

Declarations

Acknowledgements: We would like to thank Annegret Jacobs for excellent support with biosamples and Cornelia Böhme, Ela Cakmak and Jasmin Garha as Study Nurses supporting the project. Special thanks go to Leonie Wolf coordinating all administrative tasks regarding the cohort. We would also like to acknowledge Esra Özen's support with data entry into the research database. The Marga and Walter Boll Foundation provided funds to set up the MEDA-PKD infrastructure. RUM was supported by the Ministry of Science North Rhine-Westphalia (Nachwuchsgruppen.NRW 2015-2021), the German Research Foundation (DFG DI 1501/9-2, DFG MU 3629/6-1, FOR 5547/1) and the PKD Foundation. PA and RUM received support from the joergbernards-Stiftung as well as Köln Fortune and CECAD (funded by the Deutsche Forschungsgemeinschaft DFG under Germany's Excellence Strategy - EXC 2030 - 390661388). SA was supported by the Köln Fortune Program (Faculty of Medicine, University of Cologne), CECAD-Rotationsprogramm and KFH-Stiftung.

Conflicts of interest: The Dept. 2 of Internal Medicine received research funding from Otsuka Pharmaceuticals, Alnylam, Vifor and ThermoFisherScientific and the AD(H)PKD Registry was co-funded by Otsuka Pharmaceuticals. RUM served as advisor to Alnylam, AICURIS, GSK, Vertex, Vifor and is a member of the Scientific Advisory Board of Santa Barbara Nutrients. NGH receives research funding and speakers fees from Philips Healthcare, speakers fees from Elsevier and serves as advisor to Bristol Myers Squibb and BeOne Medicines.

AD(H)PKD Study Group:

Dominik Alscher, Bettina Baeßler, Hande Aydogan-Balaban, Oya Beyan, Andreas Beyer, Cornelia Böhme, Katharina Burkert, Liliana L. Caldeira, Sadrija Cukoski, Markus Cybulla, Jan Degenhardt, Marie Engelhardt, Florian Erger, Lioba Ester, Henrik ten Freyhaus, Claudia Hendrix, Astha Jaiswal, Larina Karner, Markus Ketteler, Katharina Kiefer, Adrian Kühn, Jörg Latus, Katharina Lettenmeier, Christoph Lindemann, David Maintz, Franziska Meyer, Stien Maushake, Simon Oehm, Katrin Peschel, Roman Pfister, Lena Pickert, Juliana Rank, Franz Reichel, Miriam Rinneburger, Moritz Schanz, Philipp Scherrer, Thomas Schömig, Severin Schricker, Florian Siedek, Sebastian Strubl, Victor Suarez, Janne Vehreschild, Andrea Wenzel, Fabian Wöstmann, Anna Zöll.

References

1. Ong, A. C. M., Devuyst, O., Knebelmann, B. & Walz, G. Autosomal dominant polycystic kidney disease: the changing face of clinical management. *The Lancet* **385**, 1993–2002 (2015).
2. Müller, R.-U. & Benzing, T. Management of autosomal-dominant polycystic kidney disease—state-of-the-art. *Clin Kidney J* **11**, i2–i13 (2018).
3. Spithoven, E. M. *et al.* Renal replacement therapy for autosomal dominant polycystic kidney disease (ADPKD) in Europe: prevalence and survival--an analysis of data from the ERA-EDTA Registry. *Nephrol. Dial. Transplant.* **29 Suppl 4**, iv15-25 (2014).
4. Miskulin, D. C. *et al.* Health-related quality of life in patients with autosomal dominant polycystic kidney disease and CKD stages 1-4: a cross-sectional study. *Am J Kidney Dis* **63**, 214–226 (2014).

5. Müller, R.-U. *et al.* KDIGO 2025 ADPKD guideline: a commentary on diagnosis and management of hepatopancreatic manifestations by the ERA Working Group Genes & Kidney. *Nephrol Dial Transplant* https://doi.org/10.1093/ndt/gfaf159 doi:10.1093/ndt/gfaf159.
6. Arjune, S., Todorova, P., Bartram, M. P., Grundmann, F. & Müller, R.-U. Liver Manifestations in Autosomal Dominant Polycystic Kidney Disease (ADPKD) and their impact on quality of life. *Clinical Kidney Journal* sfae363 (2024) doi:10.1093/ckj/sfae363.
7. Arjune, S. *et al.* Cardiac manifestations in patients with autosomal polycystic kidney disease (ADPKD) - a single-center study. *Kidney360* https://doi.org/10.34067/KID.0002942022 (2022) doi:10.34067/KID.0002942022.
8. Cornec-Le Gall, E., Torres, V. E. & Harris, P. C. Genetic Complexity of Autosomal Dominant Polycystic Kidney and Liver Diseases. *J. Am. Soc. Nephrol.* https://doi.org/10.1681/ASN.2017050483 (2017) doi:10.1681/ASN.2017050483.
9. Lanktree, M. B. *et al.* Intrafamilial Variability of ADPKD. *Kidney Int Rep* **4**, 995–1003 (2019).
10. Müller, R.-U. *et al.* An update on the use of tolvaptan for autosomal dominant polycystic kidney disease: consensus statement on behalf of the ERA Working Group on Inherited Kidney Disorders, the European Rare Kidney Disease Reference Network and Polycystic Kidney Disease International. *Nephrology Dialysis Transplantation* **37**, 825–839 (2022).
11. Todorova, P. *et al.* Interaction Between Determinants Governing Urine Volume in Patients With ADPKD on Tolvaptan and its Impact on Quality of Life. *Kidney International Reports* https://doi.org/10.1016/j.ekir.2023.05.011 (2023) doi:10.1016/j.ekir.2023.05.011.
12. Torres, V. E. *et al.* Tolvaptan in Patients with Autosomal Dominant Polycystic Kidney Disease. *New England Journal of Medicine* **367**, 2407–2418 (2012).
13. Torres, V. E. *et al.* Tolvaptan in Later-Stage Autosomal Dominant Polycystic Kidney Disease. *N. Engl. J. Med.* **377**, 1930–1942 (2017).
14. ClinicalSurveys.net. https://www.clinicalsurveys.net/uc/main/2890/loft/front.php?module=loft&controller=login.
15. Johnson, J. A. & Coons, S. J. Comparison of the EQ-5D and SF-12 in an adult US sample. *Qual Life Res* **7**, 155–166 (1998).
16. Aydogan Balaban, H. Ö. *et al.* Developing serum proteomics based prediction models of disease progression in ADPKD. *Nat Commun* **16**, 6646 (2025).
17. Schmidt, J. *et al.* Bridging health registry data acquisition and real-time data analytics. *Front. Med.* **11**, (2024).
18. Gehrmann, J., Herczog, E., Decker, S. & Beyan, O. What prevents us from reusing medical real-world data in research. *Sci Data* **10**, 459 (2023).
19. Schneeweiss, S. Real-World Evidence of Treatment Effects: The Useful and the Misleading. *Clin Pharmacol Ther* **106**, 43–44 (2019).

20. Steel, P. A. D., Wardi, G., Harrington, R. A. & Longhurst, C. A. Learning health system strategies in the AI era. *npj Health Syst.* **2**, 21 (2025).
21. Cornec-Le Gall, E. *et al.* The PROPKD Score: A New Algorithm to Predict Renal Survival in Autosomal Dominant Polycystic Kidney Disease. *J. Am. Soc. Nephrol.* **27**, 942–951 (2016).
22. Irazabal, M. V. *et al.* Imaging classification of autosomal dominant polycystic kidney disease: a simple model for selecting patients for clinical trials. *J. Am. Soc. Nephrol.* **26**, 160–172 (2015).
23. Ashley, E. A. The precision medicine initiative: a new national effort. *JAMA* **313**, 2119–2120 (2015).
24. El-Achkar, T. M. *et al.* Precision Medicine in Nephrology: An Integrative Framework of Multidimensional Data in the Kidney Precision Medicine Project. *American Journal of Kidney Diseases* **83**, 402–410 (2024).
25. PLD-Progression Grouper. https://halbritter-lab.github.io/pld-progression-grouper/).
26. Wolff, C. A. *et al.* Integrated Use of Autosomal Dominant Polycystic Kidney Disease Prediction Tools for Risk Prognostication. *Clin J Am Soc Nephrol* **20**, 397–409 (2025).
27. Bassanese, G. *et al.* The European Rare Kidney Disease Registry (ERKReg): objectives, design and initial results. *Orphanet J Rare Dis* **16**, 251 (2021).
28. Boerstra, B. A. *et al.* The ERA Registry Annual Report 2021: a summary. *Clin Kidney J* **17**, sfad281 (2024).
29. Wuttke, M. *et al.* A catalog of genetic loci associated with kidney function from analyses of a million individuals. *Nat Genet* **51**, 957–972 (2019).
30. Bycroft, C. *et al.* The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203–209 (2018).
31. Abaza, H. *et al.* Toward a Domain-Overarching Metadata Schema for Making Health Research Studies FAIR (Findable, Accessible, Interoperable, and Reusable): Development of the NFDI4Health Metadata Schema. *JMIR Med Inform* **13**, e63906 (2025).
32. FDPG - Forschungsdatenportal für Gesundheit. https://forschen-fuer-gesundheit.de/en/.
33. Rieke, N. *et al.* The future of digital health with federated learning. *npj Digit. Med.* **3**, 119 (2020).
34. Weiner, E. B., Dankwa-Mullan, I., Nelson, W. A. & Hassanpour, S. Ethical challenges and evolving strategies in the integration of artificial intelligence into clinical practice. *PLOS Digit Health* **4**, e0000810 (2025).
35. Pigeot, I., Intemann, T., Kollhorst, B., Sax, U. & Ahrens, W. [FAIRification of real world data for health research]. *Pravent Gesundh* 1–8 (2022) doi:10.1007/s11553-022-00973-x.
36. Evans, R. S. Electronic Health Records: Then, Now, and in the Future. *Yearb Med Inform* **Suppl 1**, S48-61 (2016).
37. Hripcsak, G. & Albers, D. J. Next-generation phenotyping of electronic health records. *J Am Med Inform Assoc* **20**, 117–121 (2013).
38. Eddy, S., Mariani, L. H. & Kretzler, M. Integrated multi-omics approaches to improve classification of chronic kidney disease. *Nat Rev Nephrol* **16**, 657–668 (2020).

39. Friedman, C. *et al.* Toward a science of learning systems: a research agenda for the high-functioning Learning Health System. *J Am Med Inform Assoc* **22**, 43–50 (2015).

Figures

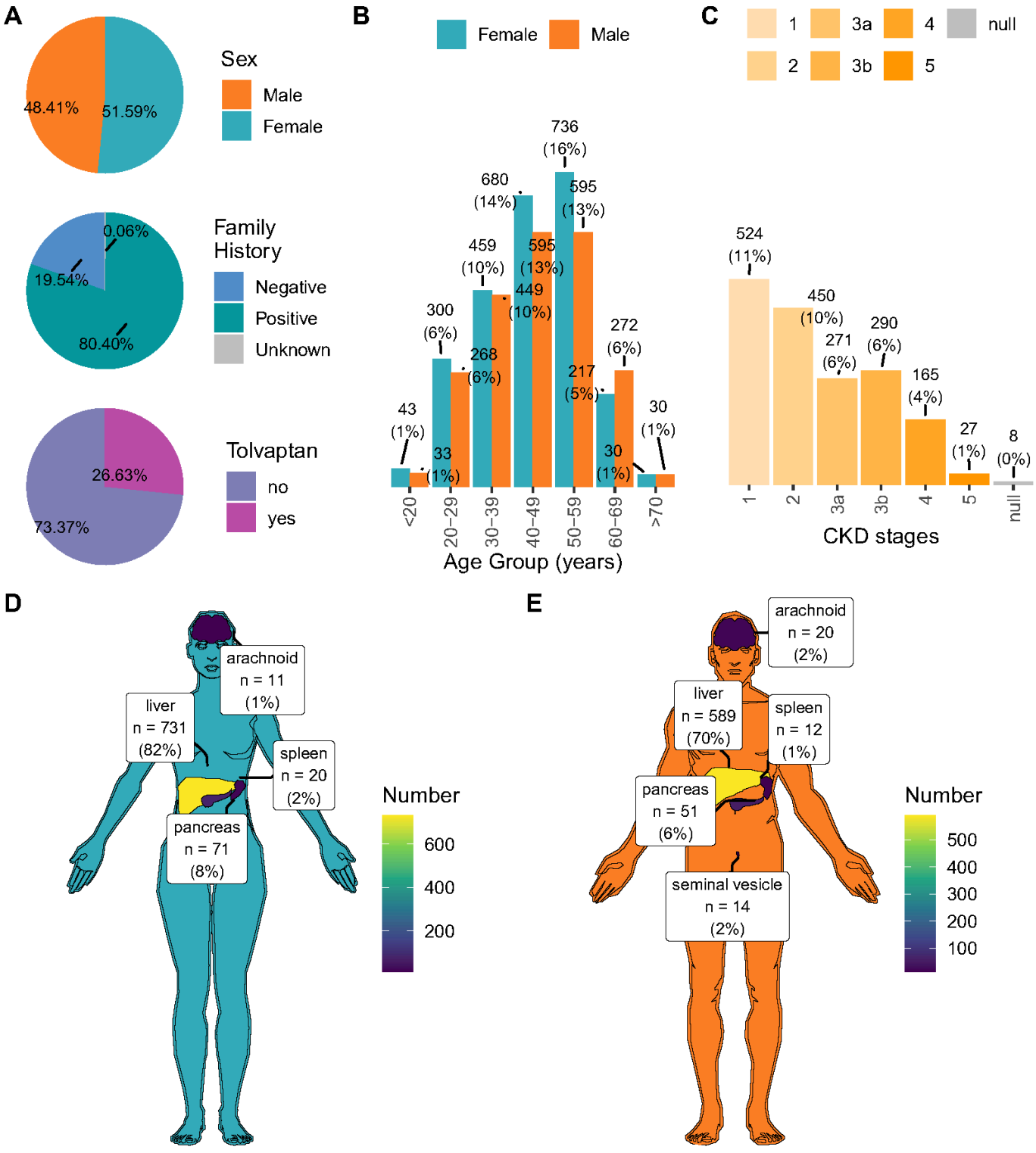


Figure 1

Dashboard “Baseline Characteristics” and „Extrarenal Manifestations“

A Fraction of total graphs showing sex, family history as well as percentage of patients on tolvaptan at any point during data collection.

B Age distribution by sex at baseline.

C CKD eGFR stages at baseline

D Extrararenal cysts in female patients

E Extrarenal cysts in male patients

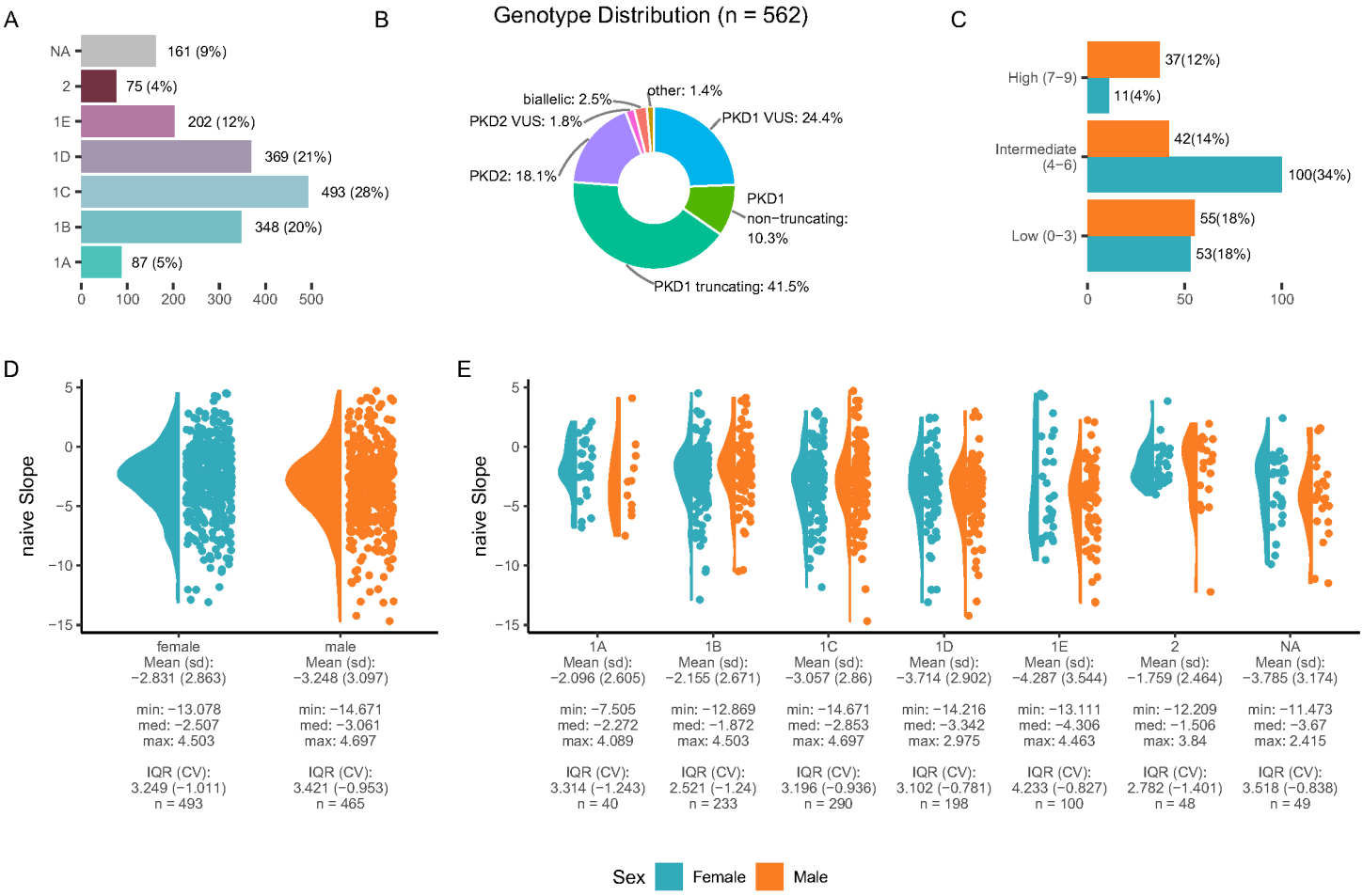


Figure 2

Dashboard “Predictors of disease progression“ and „Kidney function decline“

A Mayo Imaging Classification across the entire cohort.

B Pathogenic and likely pathogenic variants classified as PKD1 truncating, PKD1 non-truncating, PKD2 and other (genes known to be causal in ADPKD other than PKD1 and PKD2). Variants of uncertain

significance (VUS) in PKD1 and PKD2 are reported separately. Patients showing at least one pathogenic or likely pathogenic variant AND a second variant in an ADPKD-gene which is at least a VUS are reported as biallelic. (Likely) benign variants are not reported.

C PROPKD Score grouped by high, intermediated and low risk of rapid progression including data only from patients who have reached the age of 35 years and for whom genotype data is available.

For the complete dashboards see Suppl. Fig. 8 and online tool.

D eGFR slope (in ml/min/year) of male and female patients while not being on tolvaptan.

E eGFR slope (in ml/min/year) of male and female patients while not being on tolvaptan.

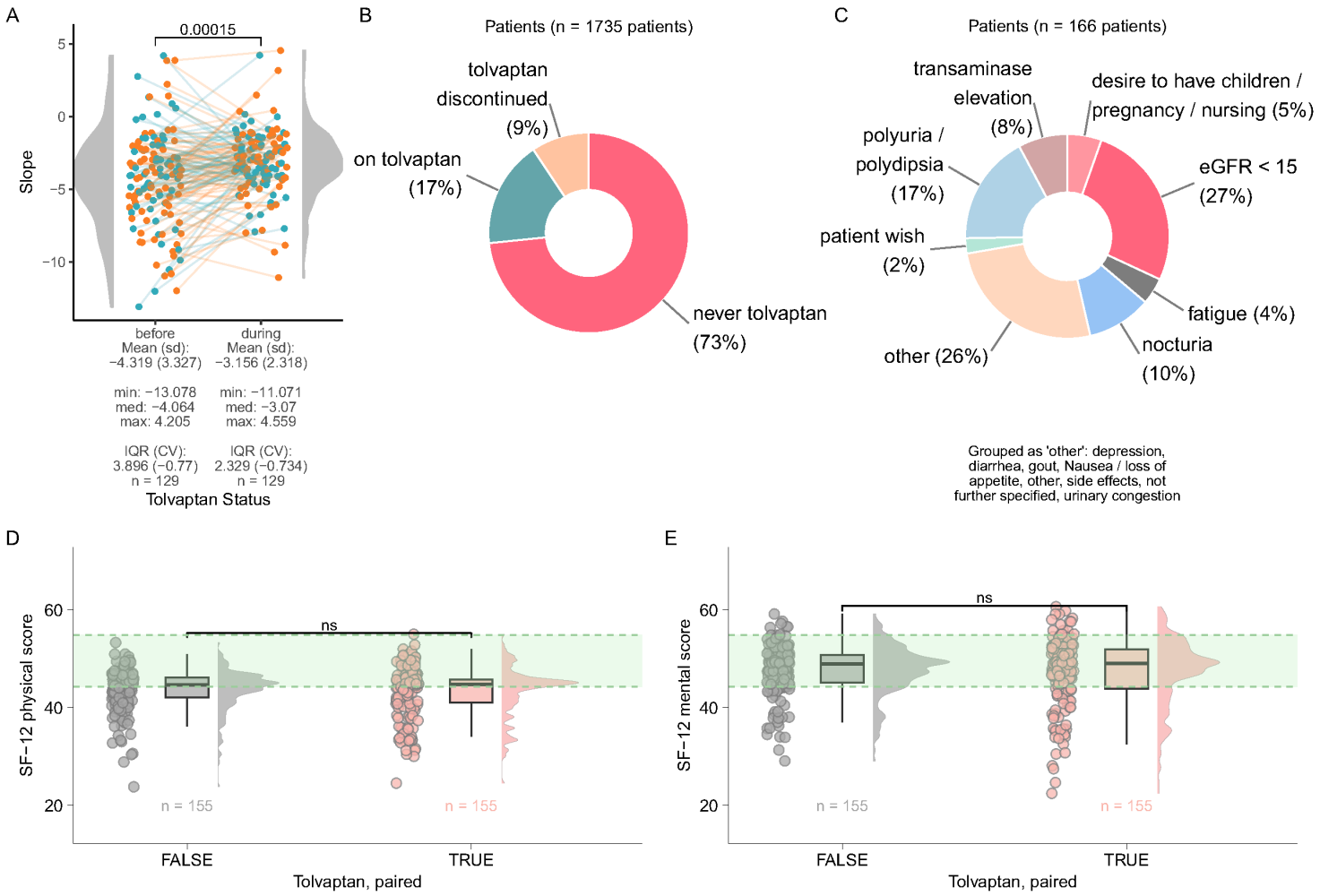


Figure 3

Dashboard "Tolvaptan" (tolvaptan usage and tolerability), effect of tolvaptan on eGFR slope (dashboard „Kidney function decline") and quality of life (dashboard „Quality of life")

A Longitudinal paired analysis of eGFR slopes before and on tolvaptan (using all eGFR values available)

B Proportion of patients who were on tolvaptan at any visit during followup across the entire cohort.

C Reasons for discontinuation of tolvaptan.

D Quality of life assessed using the SF-12 physical score in a paired analysis of patients before and on tolvaptan. Green-shaded area depicts the distribution in the normal population.

E Quality of life assessed using the SF-12 physical score in a paired analysis of patients before and on tolvaptan. Green-shaded area depicts the distribution in the normal population.

A-D: “Kidney Function Decline” Dashboard; E/F: “Quality of Life” Dashboard

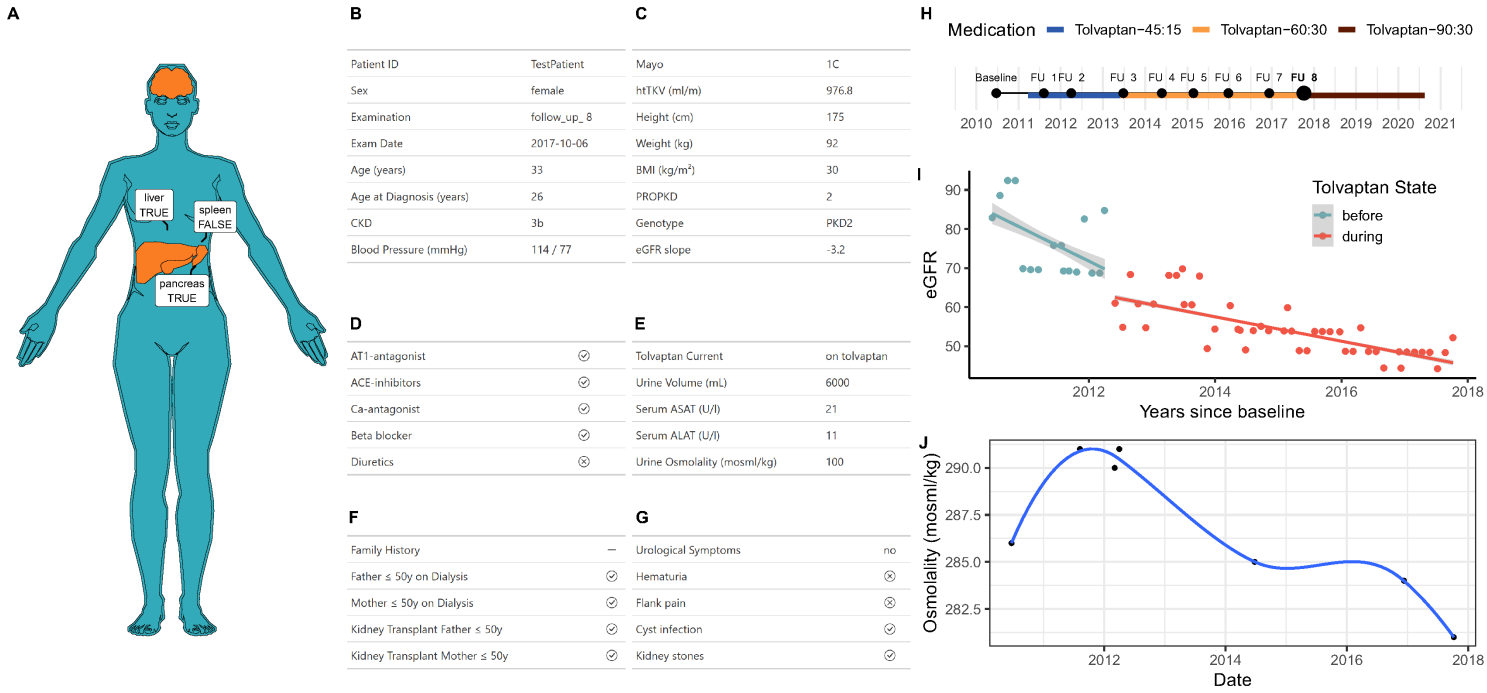


Figure 4

Individual „Patient dashboard“

A Extrarenal cysts as known at the selected visit.

B Patient characteristics at the selected visit (shown in bold) and age at diagnosis.

C Predictors of disease progression.

D Co-medication classes at the selected visit. (= yes, x = no)

E Key visit specific lab values.

F Family history (+ = positive, - = negative, = yes, x = no)

G Urological manifestations (+ = positive, - = negative, = yes, x = no)

H Visit schedule over time indicating tolvaptan treatment phases and dosage, visits can be selected by clicking upon them.

I eGFR slope indicating phases with and without tolvaptan.

J Laboratory values can be selected freely and the selected value is then visualized over time (example ALAT).

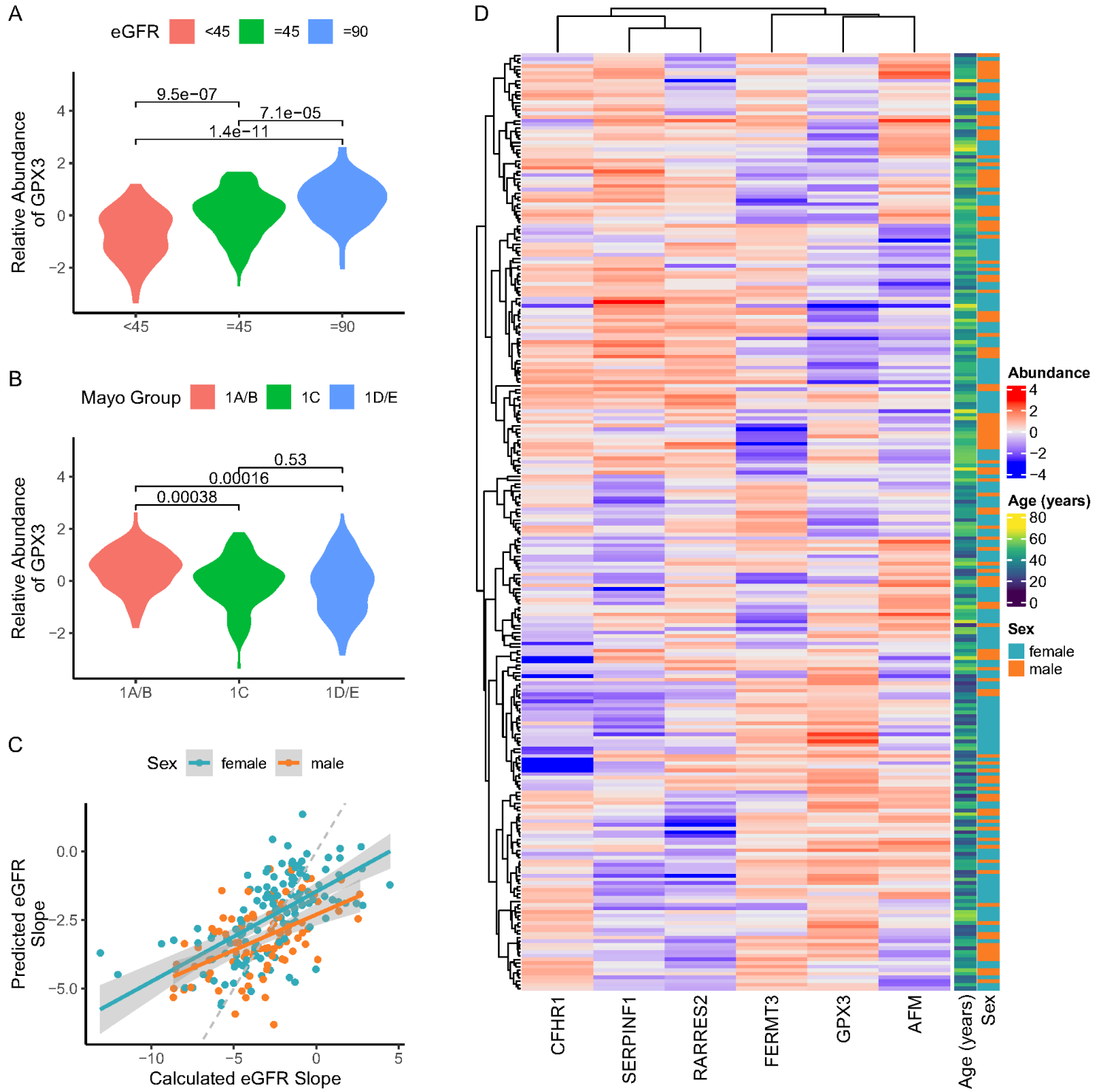


Figure 5

Integrating omics-data with clinical data to identify biomarkers for relevant outcome parameters („OMICS dashboard“)

A Relative protein abundance of proteins in the Aydogan et al. serum proteome dataset depicted by eGFR classes (GPX3 shown as an example)

B Relative protein abundance of proteins in the Aydogan et al. serum proteome dataset depicted by Mayo Imaging Classes (GPX3 shown as an example)

C Heatmap of protein abundance in conjunction with clinical parameters, both can be freely chosen by the user (showing the proteins from the Aydogan et al. 6-protein model and age / sex as examples).

D Plot showing prediction accuracy of the 6-protein model based on real-time updates of eGFR slopes.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplFigures02122025.pptx