

**Online-Only Supplementary Material**

**Table of Contents**

|                 |      |
|-----------------|------|
| Table S1 .....  | p. 2 |
| Table S2 .....  | p. 3 |
| Figure S1 ..... | p. 4 |

**Table S1. Relative importance of variables for predicting end-stage kidney disease (ESKD) using the random forest model and the Cox proportional hazard model (with missing data)**

| Random forest model           |                     |             |         |
|-------------------------------|---------------------|-------------|---------|
| C-statistic 0.935             |                     |             |         |
| Upper 12 variables            | Relative importance |             |         |
| eGFR                          | 0.088               |             |         |
| Serum albumin                 | 0.026               |             |         |
| HbA1c                         | 0.021               |             |         |
| Proteinuria, positive         | 0.014               |             |         |
| Serum bilirubin               | 0.008               |             |         |
| Serum uric acid               | 0.006               |             |         |
| Red blood cell count          | 0.005               |             |         |
| ARB use                       | 0.004               |             |         |
| Age                           | 0.003               |             |         |
| Smoking, missing              | 0.002               |             |         |
| Proteinuria, missing          | 0.001               |             |         |
| Hypertension, positive        | 0.001               |             |         |
| Cox proportional hazard model |                     |             |         |
| C-statistic 0.905             |                     |             |         |
| Upper 12 variables            | HR                  | 95% CI      | P value |
| eGFR                          | 0.240               | 0.188-0.307 | <0.001  |
| Proteinuria, positive         | 1.878               | 1.445-2.441 | <0.001  |
| HbA1c                         | 1.818               | 1.581-2.091 | <0.001  |
| Proteinuria, missing          | 1.784               | 1.289-2.470 | <0.001  |
| Serum albumin                 | 0.603               | 0.508-0.715 | <0.001  |
| Serum bilirubin               | 0.657               | 0.501-0.861 | 0.002   |
| Age                           | 0.666               | 0.554-0.800 | <0.001  |
| ARB use                       | 1.237               | 1.008-1.518 | 0.042   |
| Serum uric acid               | 1.218               | 1.014-1.463 | 0.035   |
| GLP-1R agonist use            | 0.804               | 0.665-0.973 | 0.025   |
| Gender, men                   | 1.170               | 0.974-1.405 | 0.094   |
| Hypertension, positive        | 0.903               | 0.684-1.192 | 0.472   |

\* P<0.05. ARB, angiotensin II receptor blocker; GLP-1R, glucagon-like peptide 1 receptor.

**Table S2. Characteristics of the subjects in the external validation cohort compared with those in the development cohort**

| Variables                                                     | Validation cohort | Development cohort | <i>P</i> value      |
|---------------------------------------------------------------|-------------------|--------------------|---------------------|
| n                                                             | 5153              | 2549               |                     |
| Age, years, median (IQR)                                      | 48.0(42.0-54.0)   | 57.0 (47.0-63.0)   | <0.001 <sup>a</sup> |
| Gender, male, n (%)                                           | 4622 (89.7)       | 1432 (56.2)        | <0.001 <sup>b</sup> |
| HbA1c, %, median (IQR)                                        | 6.9 (6.4-7.9)     | 7.0 (6.5-8.1)      | 0.0831 <sup>a</sup> |
| mmol/mol, median (IQR)                                        | 50.4 (45.3-60.6)  | 51.4 (46.3-62.6)   | 0.0831 <sup>a</sup> |
| Serum albumin, mg/dL, mean (SD)                               | 4.5 (0.4)         | 4.0 (0.5)          | <0.001 <sup>c</sup> |
| Serum bilirubin, mg/dL, median (IQR)                          | 0.7 (0.6-1.0)     | 0.7 (0.5-0.9)      | <0.001 <sup>a</sup> |
| eGFR, mL min <sup>-1</sup> [1.73 m] <sup>-2</sup> , mean (SD) | 80.6 (20.5)       | 81.8 (26.6)        | 0.0345 <sup>c</sup> |
| Proteinuria, n (%)                                            |                   |                    |                     |
| Positive,                                                     | 984 (19.1)        | 475 (29.7)         | 0.6492 <sup>b</sup> |
| Missing, n (%)                                                | 0                 | 948 (37.2)         |                     |
| Follow-up time, years,                                        |                   |                    |                     |
| Median (IQR)                                                  | 4.0 (3.9-4.6)     | 4.7 (2.3-8.8)      | <0.001 <sup>a</sup> |
| Mean (SD)                                                     | 4.3 (1.8)         | 5.6 (3.7)          | <0.001 <sup>c</sup> |
| ESKD, n (%)                                                   | 157 (3.1)         | 176 (6.2)          | <0.001 <sup>c</sup> |
| Time to ESKD, years,                                          |                   |                    |                     |
| Median (IQR)                                                  | 3.1 (1.6-4.5)     | 2.5 (0.9-4.8)      | <0.001 <sup>a</sup> |
| Mean (SD)                                                     | 3.4 (2.2)         | 3.2 (2.9)          | <0.001 <sup>c</sup> |

a Calculated by Mann Whitney test, b Calculated by c2 test, c Calculated by t test. ESKD, end-stage kidney disease. Abbreviation: eGFR, estimated glomerular filtration rate; HbA1c, hemoglobin A1c.

