

The 67th ASH Annual Meeting Abstracts

ORAL

652. MGUS, AMYLOIDOSIS, AND OTHER NON-MYELOMA PLASMA CELL DYSCRASIAS: CLINICAL AND EPIDEMIOLOGICAL

AI-based analysis reveals novel blood protein biomarkers for predicting cardiac amyloidosis approximately a decade early: A UK biobank study

Seyed Alireza Hasheminasab¹, Berrin Balik¹, Joshua Fieggen², Lei Clifton³, Nikant Sabharwal⁴, Karthik Ramasamy^{1,5}, Ashutosh Wechalekar⁶, Anjan Thakurta¹

¹Oxford Translational Myeloma Centre (OTMC), Nuffield Department of Orthopedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, Oxford, UK, Oxford, United Kingdom

²Computational Health Informatics (CHI) Lab, Institute of Biomedical Engineering, Department of Engineering Sciences, University of Oxford, Oxford, United Kingdom

³Nuffield Department of Primary Care Health Sciences, University of Oxford, Oxford, United Kingdom

⁴Oxford University Hospitals NHS Foundation Trust, Cardiology, Oxford, United Kingdom

⁵Oxford University Hospitals NHS Foundation Trust, Haematology, Oxford, United Kingdom

⁶National Amyloidosis Centre, University College London, London, United Kingdom

Abstract Background: Amyloidosis are a complex debilitating group of diseases characterised by the extracellular deposition of abnormal amyloid fibrils, leading to progressive organ dysfunction. With limited reversibility of end organ failure, accurate early prediction of disease onset before clinically apparent symptoms is critical for timely intervention and improved outcomes but remains challenging. We report here use of novel, non-invasive biomarkers by large-scale proteomic profiling as promising approach for early identification.

Aims: To identify and characterize protein biomarkers associated with amyloidosis using statistical and machine learning approaches from the UK Biobank (as serum repository n=2,923), and to initiate cross-platform validation of key candidates in real-world patient samples.

Methods: Of 502,128 total participants in the UK Biobank we identified 422 individuals (0.08%) with amyloidosis. We restricted detail analysis to 53,013 participants with full proteomics data of which 47,849 had data linked hospital records. After excluding individuals with a diagnosis of amyloidosis prior to the index date (defined as the date of protein assessment), we identified 47,782 controls and 61 cases (0.1%) of amyloidosis with proteomics data. Compared to the overall UK Biobank proteomics population, the amyloidosis group was older (mean age: 62.75 years; standard deviation: 6.04), while gender distribution was broadly similar, with a slightly higher proportion of males in the disease group. When compared to 422 amyloidosis patients in UK Biobank without proteomics data, the cases with proteomics demonstrated a similar age and gender distribution, supporting the representativeness of this subgroup within the broader amyloidosis population and, notably, similar to the demographic profile reported in global epidemiological trends in 2022.

We conducted univariate differential protein expression analysis as well as a time-to-diagnosis prediction analysis. The differential expression analysis was adjusted for age and sex to identify proteins significantly associated with amyloidosis. This analysis was performed using the Limma package, which applies empirical Bayes methods to improve statistical power, particularly in settings with small and imbalanced sample sizes. Additionally, we developed a time-to-event prediction model using XGBoost, optimised with a Cox proportional hazards loss function. To identify key proteomic predictors of amyloidosis progression, we applied SHapley Additive exPlanations (SHAP). Ongoing cross-platform validation of these biomarkers is being initiated in a prospective cohort of AL and ATTR patients.

Results: Univariate analysis identified specific protein alterations associated with cardiac functions and amyloidosis. A time-to-event prediction model, trained on 70% of the cohort, with the remaining 30% as “held-out” test dataset, demonstrated strong performance with a concordance index (C-index) of 0.81 on the test dataset. The model identified 15 proteins, partially overlapping with those from the univariate analysis, suggesting potential interactions among proteins and highlighting the importance of temporal dynamics in disease prediction. Time-dependent area under the receiver operating characteristic curve (AUROC) values on the test dataset were 0.7958, 0.7952, 0.7863, 0.8035, and 0.8086 at 4, 6, 8, 10, and 12 years from the

baseline assessment, respectively. Several proteins identified in the univariate analysis, along with additional ones revealed by the XGBoost model, emerged as important predictors of disease risk and progression. We are currently initiating prospective validation studies in patient samples with AL and ATTR amyloidosis to investigate the role of these proteins in disease progression and subtype classification. Full analysis of the proteomic data will be presented.

Conclusion: This study identified novel abnormally elevated protein biomarkers about a decade before clinical diagnosis of amyloidosis using a large population-based cohort, applying both univariate statistical and multivariate machine learning approaches, consistent with published clinical observations of “asymptomatic” amyloid deposits in historical tissue biopsies. The strong predictive performance of our model, together with ongoing cross-platform validation in real-world patient samples, supports further investigation of these biomarkers for their potential utility in early prediction of amyloidosis.

<https://doi.org/10.1182/blood-2025-580>