

# Hub Genes Identification, Small Molecule Compounds Prediction for Atrial Fibrillation and Diagnostic Model Construction Based on XGBoost Algorithm



Lingzhi Yang<sup>1</sup>, Wei Huang<sup>1</sup>

<sup>1</sup>Department of Cardiology, the First Affiliated Hospital of Chongqing Medical University, China

# 1. INTRODUCTION

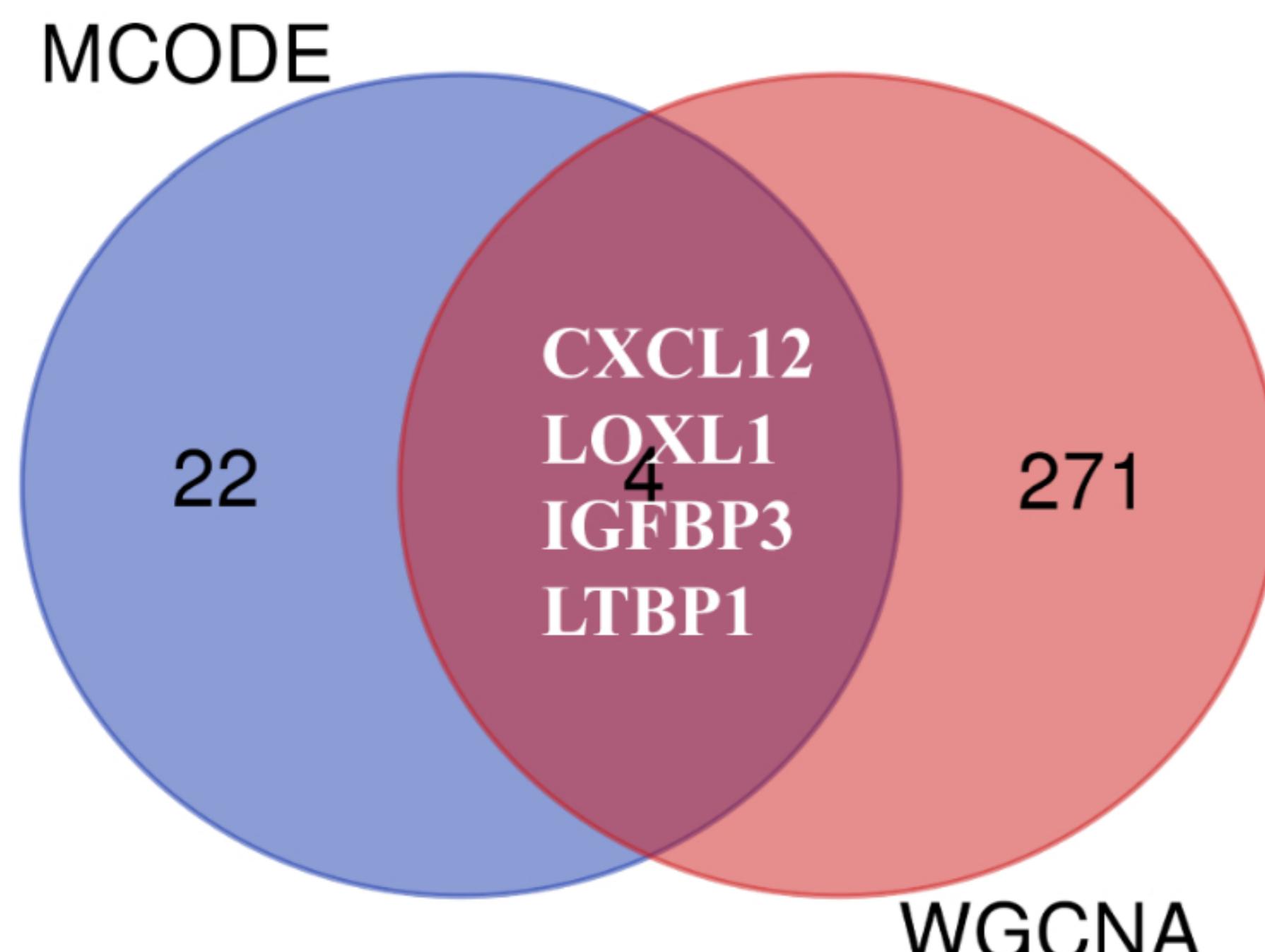
Atrial fibrillation (AF) is the most common sustained cardiac arrhythmia and engenders significant global health care burden. The underlying mechanisms of AF is remained to be revealed and current treatment options for AF have limitations. Besides, a detection system can help identify those at risk of developing AF and will enable personalized management.

## 2. METHODS AND MATERIALS

Six AF microarray datasets were used for transcriptomic analyses based on standard pipelines to identify key genes, pathways related to AF. Hub gene CXCL12 was analyzed comprehensively for subsetting AF. Machine Learning-based algorithm XGBoost was used for diagnostic model construction.

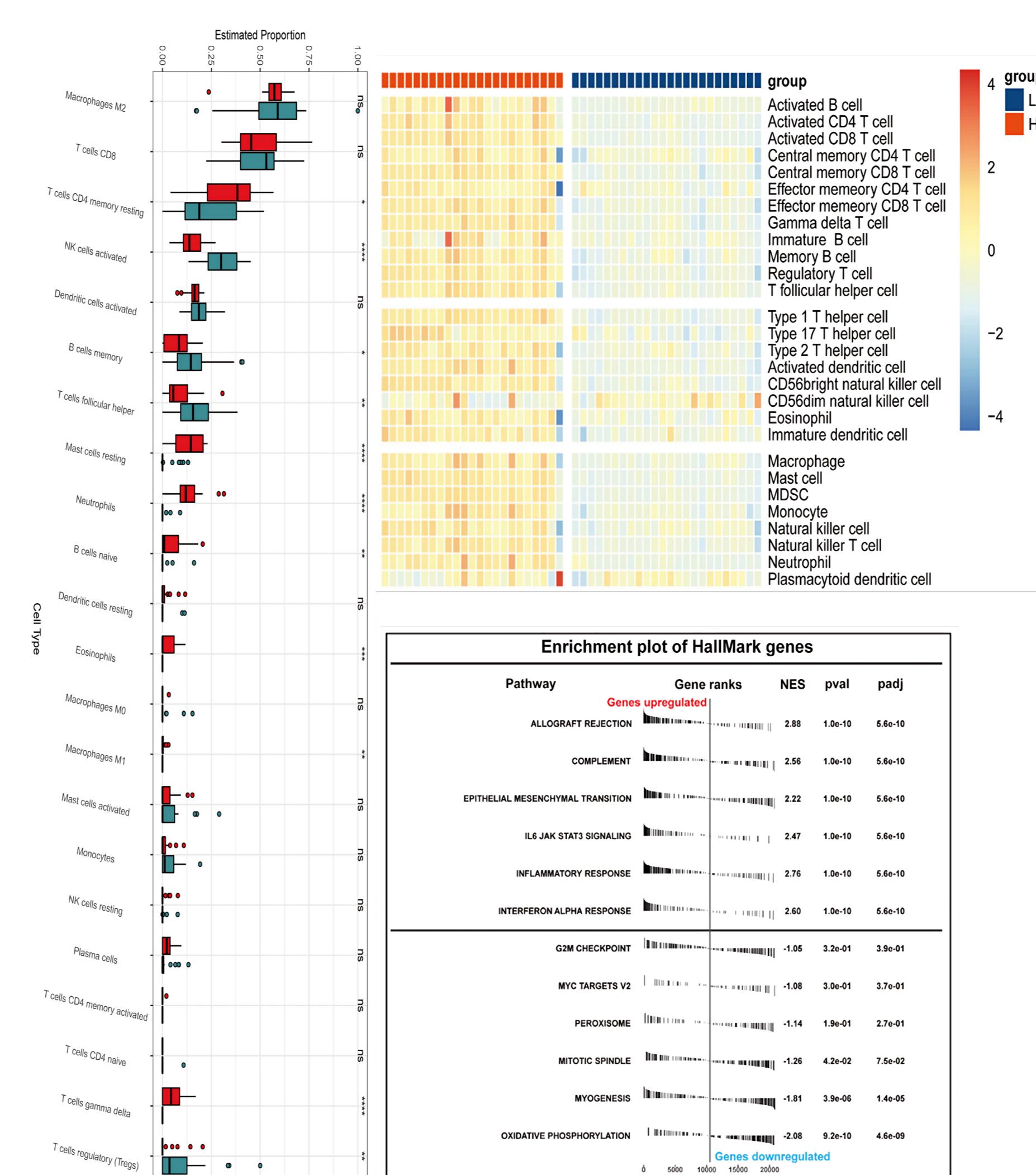
## 3. RESULTS

Identification of candidate genes for AF. By combining Robust Rank aggregation (RRA) method, Weighted Gene Co-expression Network Analysis (WGCNA) and Molecular Complex Detection (MCODE) using protein-to-protein network, we identified CXCL12, LTBP1, LOXL1, and IGFBP3 as candidate genes.



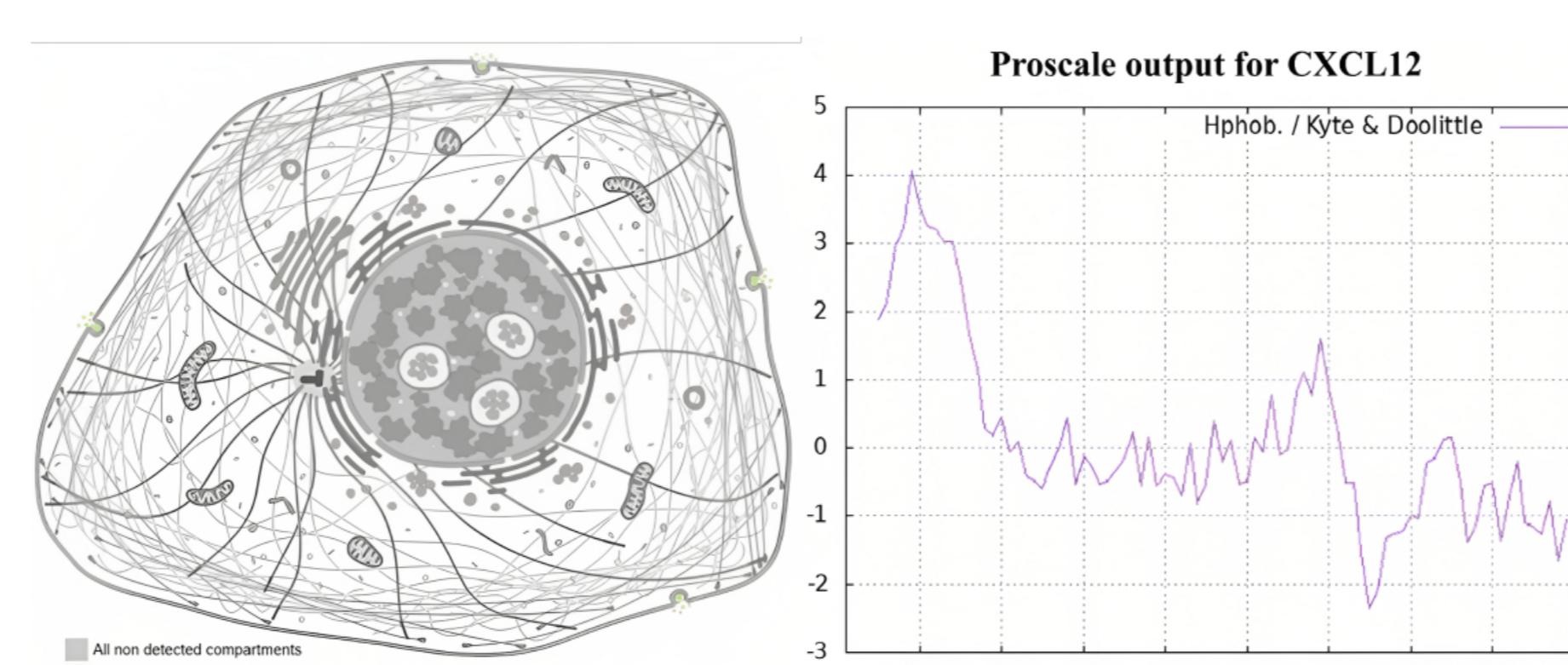
**Figure 1:** Venn plot showing the overlapped genes

CD4 memory T cells, mast cells, neutrophils, and gamma delta ( $\gamma\delta$ ) T cells showed greater infiltration in the high CXCL12 group, and Treg cells showed lower levels of infiltration in the high CXCL12 group. More inflammatory and immune response pathways were enriched in high CXCL12 group.



**Figure 2:** Immune infiltration analysis between high and low CXCL12 group in AF and functional enrichment analysis

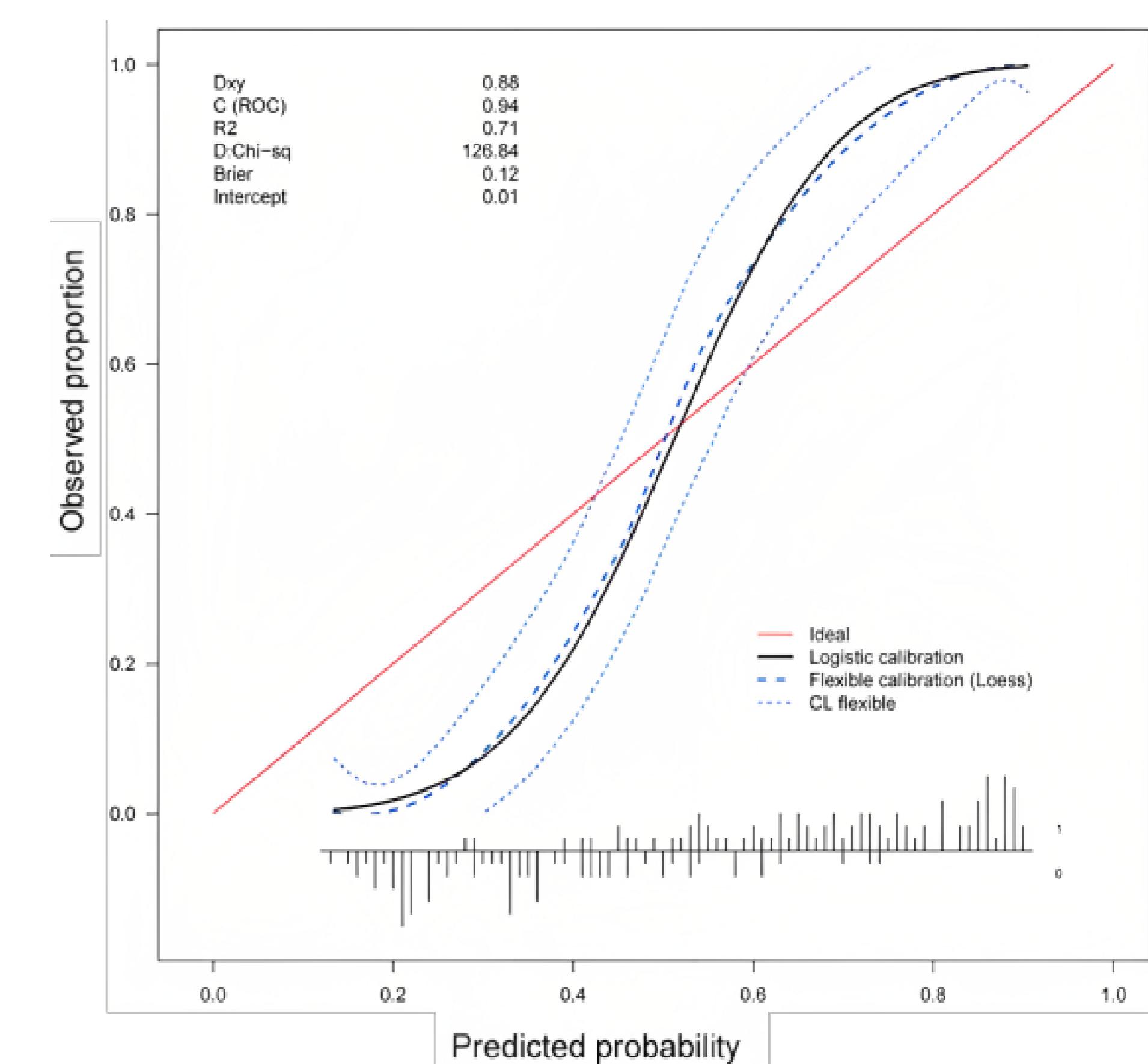
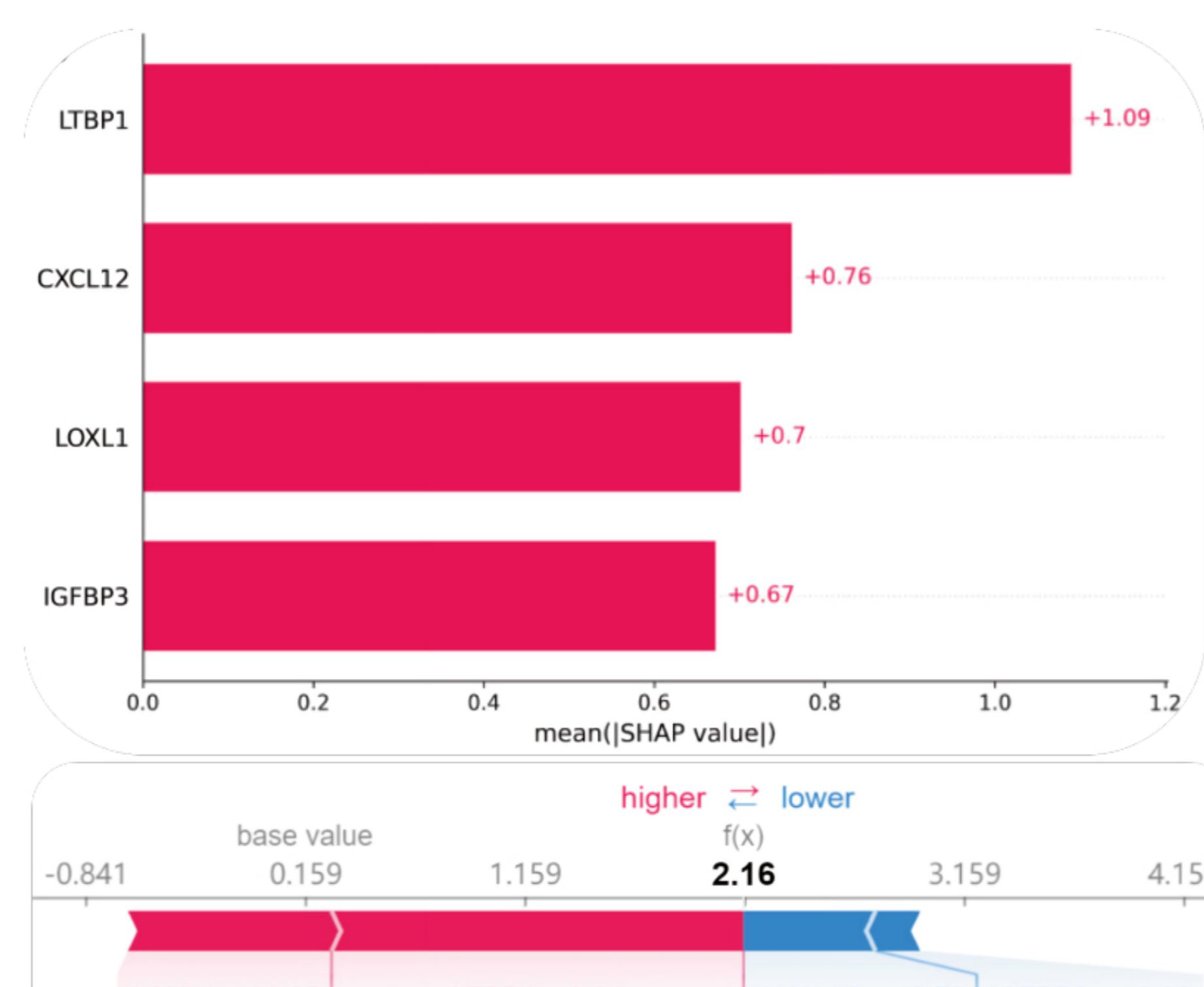
CXCL12 is a secretary, stable protein with both hydrophobic and hydrophilic properties.



**Figure 3.** Biological features of CXCL12

The diagnostic model constructed by XGBoost algorithm shows great performance with an AUC of 0.9385 (95 CI: 0.9044–0.9725) and brier score of 0.12.

The variance importance plot showed the importance of included variables and the force plot showed an individual case.



The figure is a Receiver Operating Characteristic (ROC) plot. The vertical axis is labeled "True Positive Rate (Sensitivity)" and ranges from 0.00 to 1.00. The horizontal axis is labeled "False Positive Rate (1 - Specificity)" and ranges from 0.00 to 1.00. A solid black step-line represents the model's performance, starting at (0.0, 0.35) and ending at (1.0, 1.00). A light blue shaded area surrounds this line, representing the 95% confidence interval. A dashed diagonal line from (0.0, 0.0) to (1.0, 1.0) serves as a reference for a random classifier. Text at the bottom right of the plot area states "AUC of XGBoost: 0.9385(95% CI: 0.9044–0.9725)".

**Figure 5:** Diagnostic model construction-validation

# 4. CONCLUSIONS

Four key genes involved in the pathogenesis of AF. The biological features of CXCL12 showed its potential implication as a marker to distinguish AF subsets, and showed that it could be an important intermediate between the local inflammatory microenvironment and atrial fibrosis

## **5. FORTHCOMING RESEARCH**

- The current study only included bioinformatics analyses. Future *in vitro* and *in vivo* studies will be needed to explore the molecular mechanisms and pathways identified in this study.
  - Detection CXCL12 might help identify AF subsets at molecular level