**Target Score: a proteomics data selection tool applied to esophageal cancer identifies GLUT1-Sialyl Tn glycoforms as biomarkers of cancer aggressiveness**

Sofia Cotton1,2,3,4, Dylan Ferreira1,2,3,4, Janine Soares1,2,5,Andreia Peixoto1,3,4, Marta Relvas-Santos1,2,3,4,6, Rita Azevedo7, Paulina Piairo8, Lorena Diéguez8, Carlos Palmeira1,9,10, Luís Lima1,11, André M.N. Silva6, Lúcio Lara Santos1,2,10,11,12, José Alexandre Ferreira1,2,11

1Experimental Pathology and Therapeutics Group, IPO Porto Research Center (CI-IPOP), Portuguese Oncology Institute (IPO Porto), 4200-072 Porto, Portugal; 2Institute of Biomedical Sciences Abel Salazar (ICBAS), University of Porto, 4050-313 Porto, Portugal; 3Institute for Research and Innovation in Health (i3S), University of Porto, 4200-135 Porto, Portugal; 4Institute for Biomedical Engineering (INEB), Porto, Portugal, 4200-135 Porto, Portugal; 5QOPNA/LAQV, Department of Chemistry, University of Aveiro, Campus Universitário de Santiago, 3810-193, Aveiro, Portugal; 6REQUIMTE-LAQV, Department of Chemistry and Biochemistry, Faculty of Sciences of the University of Porto, 4169-007 Porto, Portugal; 7Institute of Biomedicine, University of Turku, FI-20014 Turku, Finland; 8Medical Devices Research Group, International Iberian Nanotechnology Laboratory (INL), 4715-330 Braga, Portugal; 9Departament of Immunology, Portuguese Institute of Oncology of Porto, 4200-072 Porto, Portugal; 10Health School of Fernando Pessoa University, 4249-004 Porto, Portugal; 11Porto Comprehensive Cancer Center (P.ccc), 4200-072 Porto, Portugal; 12Department of Surgical Oncology, Portuguese Institute of Oncology, 4200-072 Porto, Portugal;

**Corresponding author:**

José Alexandre Ferreira (jose.a.ferreira@ipoporto.min-saude.pt)

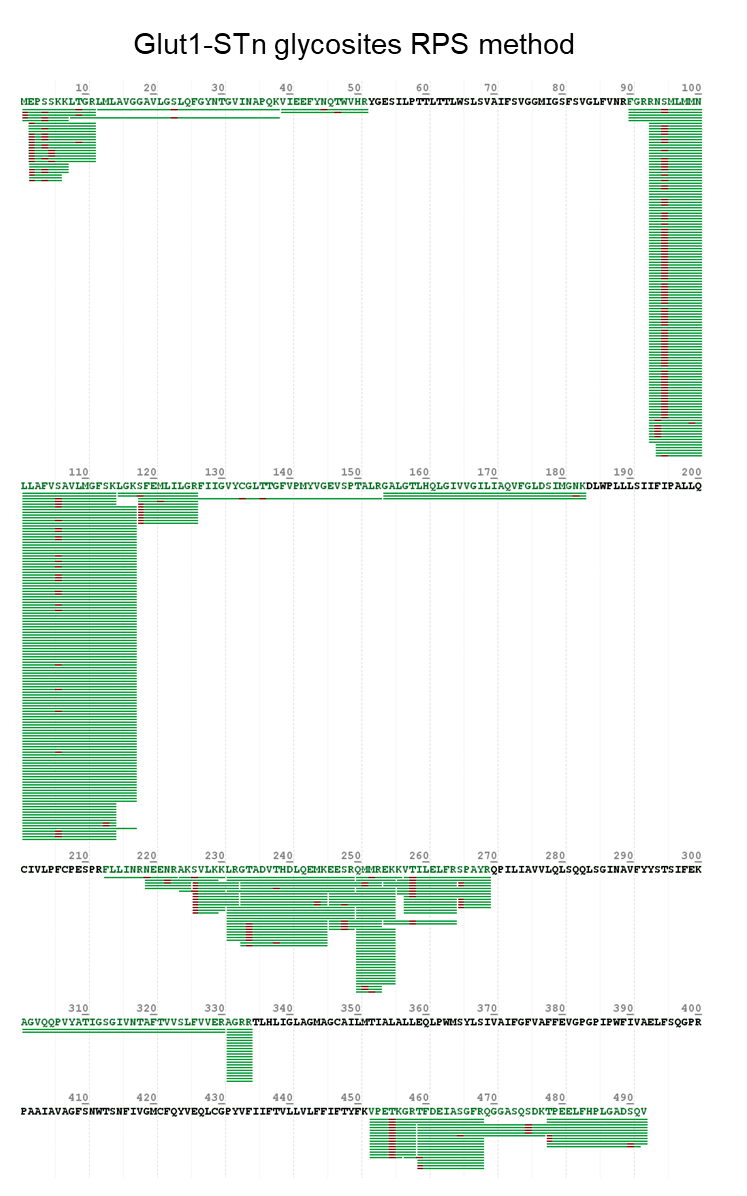
Experimental Pathology and Therapeutics Group, Research Centre, Portuguese Oncology Institute of Porto, R. Dr. António Bernardino de Almeida 62, 4200-072 Porto, Portugal; Tel. +351 225084000 (ext. 5111).

**Running head:** Targetable glycoproteins in esophageal cancer

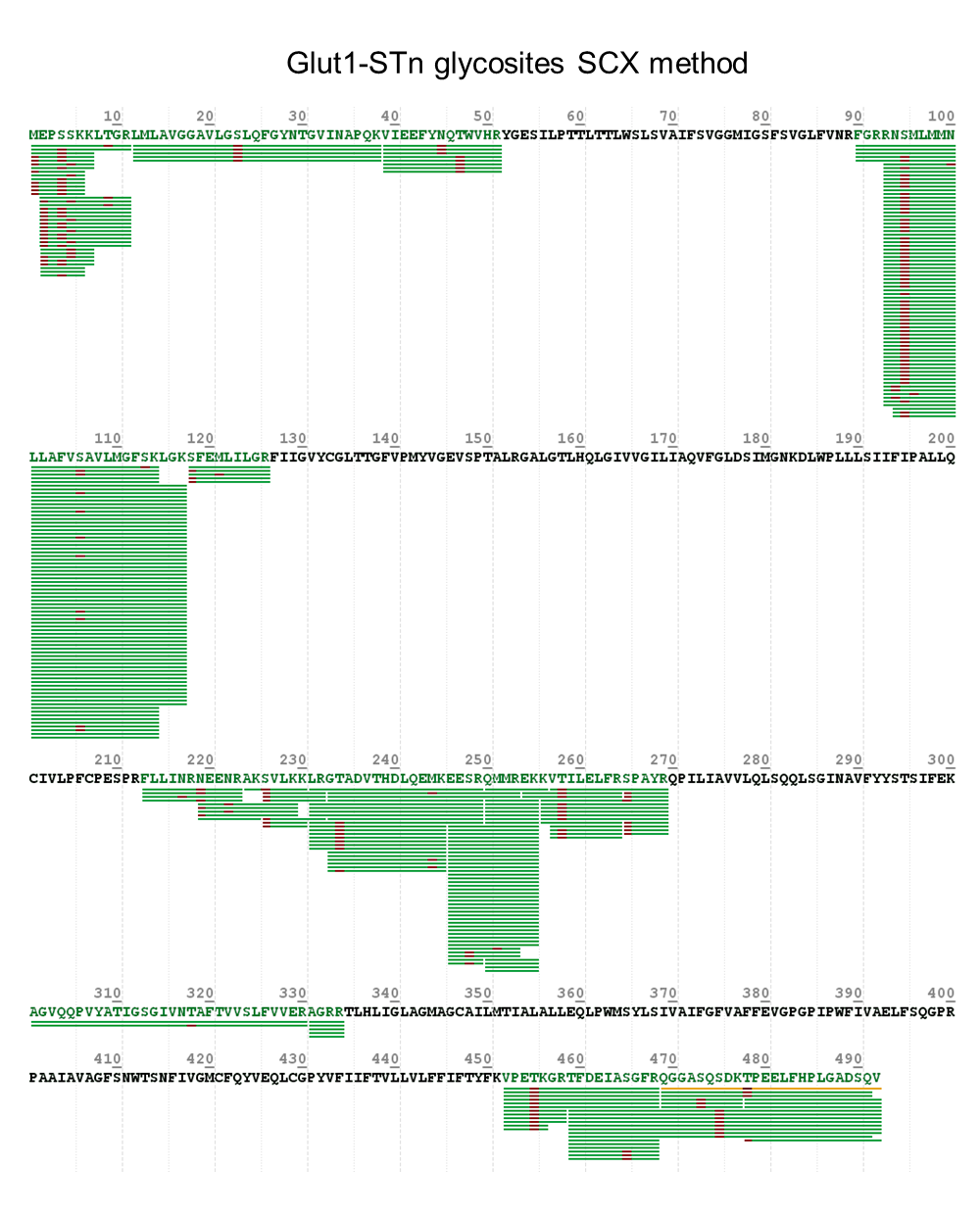
**Keywords:** cancer biomarkers; circulating tumours cells; esophageal cancer; glycomics; glycoproteomics; bioinformatics

**Supporting Material and Methods**

Proteomics data generates a high number of protein identifications, warranting tools that allow to easily pinpoint biomarkers of potential clinical interest. The Target Score algorithm was developed to provide an easy and comprehensive navigation throughout pre-existing molecular databases, enabling to extract information that maybe used to rank identified biomarkers in relation to prognosis and potential targetability. It was designed to attribute higher scores to glycoproteins overexpressed in cancer, located at the cell membrane, and associated with poor prognosis, while penalizing glycoprotein remaining at the same subcellular location than previously observed in healthy tissues. As primary source of information for this study, we used the Protein Atlas database (https://www.proteinatlas.org/) consulted in March 2020. Nevertheless, it sets a roadmap that maybe used to navigate through other databases containing information on protein expression in cancer and healthy tissues, their subcellular location and cancer prognosis. Accordingly, the score system results from the sum of the following nine variables: i) location in healthy cells (membrane: 0 points; other subcellular locations: 1 point); ii) location in cancer cells (membrane: 1 point; cytoplasm and/or other subcellular locations: 0 points); iii) expression in healthy esophageal epithelium at the plasma membrane (negative: 3 points; low: 2 points; moderate: 1 point; high: 0 points); iv) expression in head-and-neck cancer (negative: 0 points; low: 1 point; moderate: 2 points; high: 3 points); v) prognosis value in cancer (high expression is not prognostic and/or associates with favorable prognosis: 0 points; high expression associates with poor prognosis: 1 point); vi) expression in lymphoid tissues at the plasma membrane (not expressed: 1 point; expressed: 0 points); vii) expression in gametes (not expressed: 1 point; expressed: 0 points); viii) index of expression in healthy tissues at the plasma membrane (varies from 3 points (no expression) to 0 (high expression)).



**Figure S1. GLUT1-STn glycopeptides (green) and glycosites (red) identified in ESCC using the RPS method.** Assignments were made using the Byonic software.

****

**Figure S2. GLUT1-STn glycopeptides (green) and glycosites (red) identified in ESCC using the SCX method.** Assignments were made using the Byonic software.