Analysis of mutation-originated gain-of-glycosylation using mass spectrometry-based N-glycoproteomics

ZHIXIN TIAN¹ and Hailun Yang¹

¹Tongji University

October 21, 2023

Abstract

RATIONALE: A general N-glycoproteomics analysis pipeline has been established for characterization of mutation-related gain-of-glycosylation (GoG) at intact N-glycopeptide molecular level, generating comprehensive site and structure information of N-glycosylation. METHODS: This study focuses on mutation-originated gain-of-glycosylation using mass spectrometry-based N-glycoproteomics analysis workflow. In brief, GoG intact N-glycopeptide databases were built, consisting of 2,701 proteins (potential GoG N-glycosites and amino acids derived from MUTAGEN, VARIANT and VAR_SEQ in UniProt) and 6,709 human N-glycans ([?]50 sequence isomers per monosaccharide composition). We employed the site- and structure-specific N-glycoproteomics workflow utilizing intact N-glycopeptides search engine GPSeeker to identify GoG intact N-glycopeptides from parental breast cancer stem cells (MCF-7 CSCs) and adriamycin-resistant breast cancer stem cells (MCF-7/ADR CSCs). RESULTS: With the criteria of spectrum-level FDR control of [?]1%, we identified 88 and 96 GoG intact N-glycopeptides corresponding to 38 and 36 intact N-glycoproteins from MCF-7 CSCs and MCF-7/ADR CSCs, respectively. Among KEGG annotation of GoG N-glycoproteins, DNA polymerase eta (POLH), serine-protein kinase ATM (ATM) and cellular tumor antigen p53 (P53) were enriched in platinum drug resistance signal pathway. ATM, P53 and G2 and S phase-expressed protein 1 (GTSE1) were associated with p53 signaling pathway. CONCLUSIONS: The integration of site- and structure-specific N-glycoproteomics approach, conjugating with GoG characterization, provides a universal workflow for revealing comprehensive N-glycosite and N-glycan structure information of gain-of-glycosylation. The analysis of mutation-originated gain-of-glycosylation can be extended GoG characterization to all the other N-glycoproteome systems including complex clinical tissues and body fluids.

Hosted file

(A) Q9Y215
VAR_SEQ

SPVTVVACNESQACLLPR
N269

(B) Q9HCH5
VAR_SEQ

KYTYQLPGNESSK
N117

VARIANT

LNTQNATAFR
N335

VAR_SEQ

SNGLESQVNQCDK
N666