Supporting Information for Publication

A quantitative longitudinal inventory of the *N*-glycoproteome of human milk from a single donor reveals the highly variable repertoire and dynamic site-specific changes

Jing Zhu[a,b,c,#], Yu-Hsien Lin[a,b,#], Kelly A. Dingess[a, b], Marko Mank[d], Bernd Stahl[d, e] and Albert J. R. Heck[a,b*]

[a] Biomolecular Mass Spectrometry and Proteomics, Bijvoet Center for Biomolecular Research and Utrecht Institute for Pharmaceutical Sciences, University of Utrecht, Padualaan 8, 3584 CH Utrecht, The Netherlands

[b] Netherlands Proteomics Center, Padualaan 8, 3584 CH Utrecht, The Netherlands

[c] Beijing Institute of Nutritional Resources, 100069, Beijing, China

[d] Danone Nutricia Research, Uppsalalaan 12, 3584 CT Utrecht, The Netherlands

[e] Chemical Biology & Drug Discovery, Utrecht Institute for Pharmaceutical Sciences University of Utrecht, The Netherlands

[#] Equal Contribution

* Authors for Correspondence: Albert Heck, a.j.r.heck@uu.nl (+31302536797)

Figure S4. Visual summary for each identified glycoprotein. Schematic drawing of the sequences of the identified glycoproteins with identified sites and identified different glycan compositions. (provided as separate file)

Table S1. 200 possible N-glycan compositions in human without sodium and with multiple fucoses

Table S2. Peptide to spectrum matches for EThcD spectra

Table S3. Identified N-glycopeptides in human milk across lactation

Table S4. Glycan compositions at identified N-glycosites

Table S5. The identities of the outer nodes in the glycoprotein-glycan network diagram (Figure 2C)

Table S6. The normalized intensity of glycopeptides across lactation in one single donor

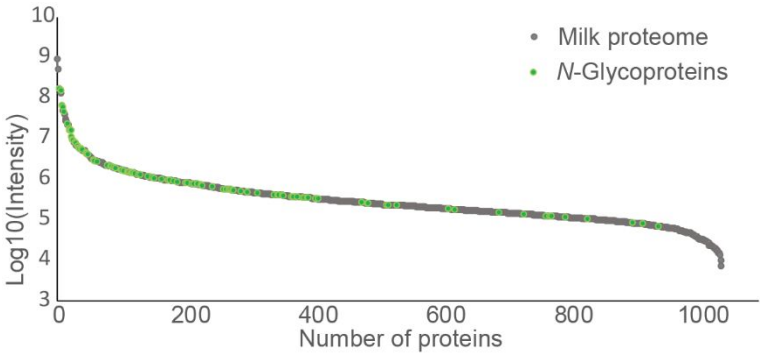

Figure S1. Dynamic range in the human milk proteome and identified glycoproteins. The dynamic range of detected glycoproteins spans the whole range of proteins detected in the human milk proteome. Proteins detected in the human milk proteome are plotted by their intensity in grey and the glycoproteins identified are mapped on top of them and colored in green.

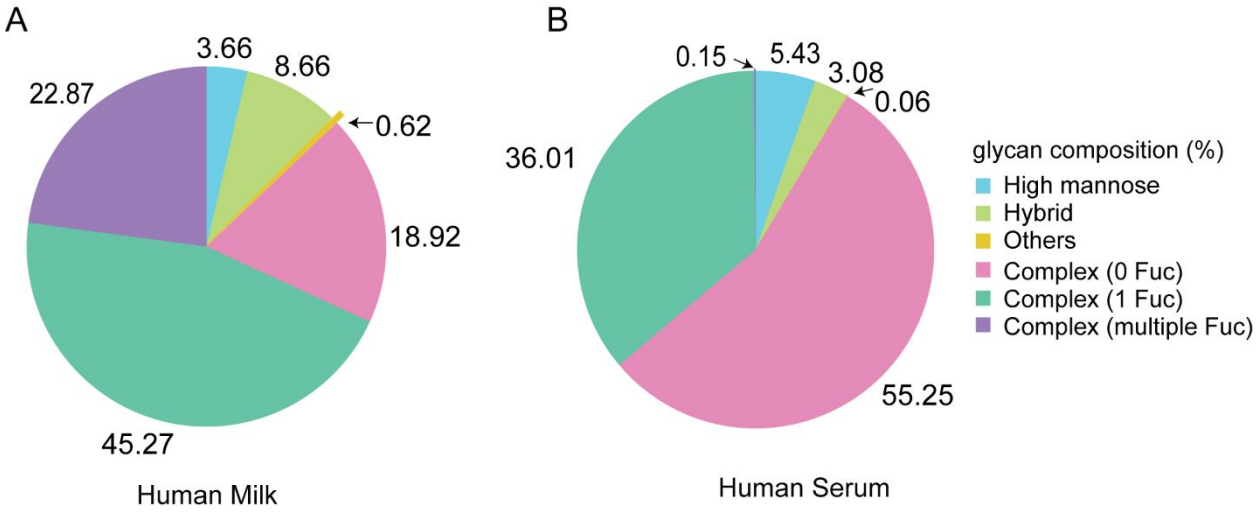

Figure S2. Global comparison of the glycan composition of *N*-glycopeptides in human milk vs human serum. A. Composition of N-linked glycoforms by PSMs in human milk based on PSMs (this study). B. Composition of N-linked glycoforms by PSMs in human serum based on data from Sun *et al.*[23]. In particular, hybrid structures and complex glycans harboring multiple fucoses seems to be more abundant in human milk vs serum.

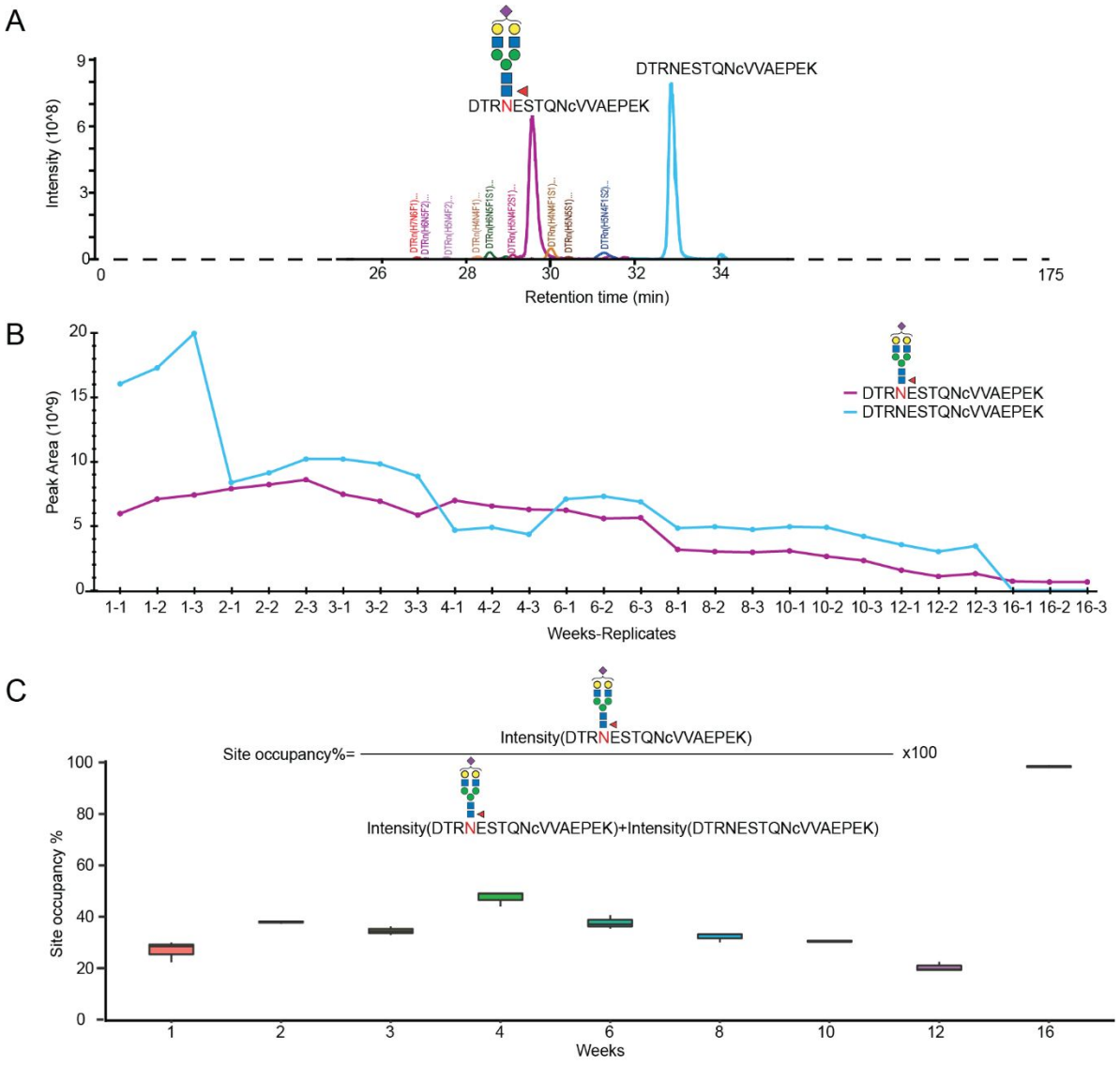

Figure S3. Represented N-glycan site occupancy of Asn69 on alpha-S1-casein as extracted from data from one major abundant glycopeptide and the unmodified peptide. A. An illustrative example of XICs of the unmodified peptide DTRNESTQNCVVAEPEK and its glycosylated form harboring a HexNAc(4)Hex(5)Fuc(1)Sia(1) glycan and other less abundant glycosylated forms. B. The XICs of the unmodified peptide DTRNESTQNCVVAEPEK and its glycosylated form harboring a HexNAc(4)Hex(5)Fuc(1)Sia(1) glycan in all replicates. C. Observed site occupancy across the stages of lactation measured in triplicates. The glycosylation site occupancies ranged from 20% (weeks 1 and 12) to almost fully occupied 95% (week 16).

```
SP|P47710-1|CASA1_HUMAN MRLLILTCLVAVALARPKLPLRYPERLQNPSE-----------------SSE-PIPLESR 42
SP|P47710-2|CASA1_HUMAN MRLLILTCLVAVALARPKLPLRYPERLQNPSE-----------------SSE-PIPLESR 42
SP|P47710-3|CASA1_HUMAN MRLLILTCLVAVALARPKLPLRYPERLQNPSE-----------------SSE-PIPLESR 42
SP|P47710-4|CASA1_HUMAN MRLLILTCLVAVALARPKLPLRYPERLQNPSE-----------------SSE-PIPLESR 42
SP|P02662|CASA1_BOVIN   MKLLILTCLVAVALARPKHPIKHQGLPQE-----------VLN-ENLLRFFVAPFPEVFG 48
SP|P19228|CASA1_MOUSE   MKLLILTCLVAAAFAMPRLHSRNAVSSQTQQQHSSS--EEIFKQ---------PKYLNLN 49
SP|P04653|CASA1_SHEEP   MKLLILTCLVAVALARPKHPIKHQGLSSE-----------VLN-ENLLRFVVAPFPEVFR 48
SP|P02661|CASA1_RAT     MKLLILTCLVAAALALPRAHRRNAVSSQTQQENSSSEEQEIVKQ---------PKYLSLN 51
SP|P18626|CASA1_CAPHI   MKLLILTCLVAVALARPKHPINHRGLSPE-----------VPN-ENLLRFVVAPFPEVFR 48
SP|O97943|CASA1_CAMDR   MKLLILTCLVAVALARPKYPLRYPEVFQNEPDSIE----EVLNKRKILELAVVSP-IQFR 55
SP|P39035|CASA1_PIG     MKLLIFICLAAVALARPKPPLRHQEHLQNEPDSRE----ELFKERKFLRFPEVPLLSQFR 56
SP|P86272|CASA1_EQUAS   ---------------RPKLPHRHPEIIQNEQDSRE----KVLKERKFPSFALHTPRE--- 38
SP|P09115|CASA1_RABIT   MKLLILTCLVATALARHKFHLGHLKLTQEQPESSE---QEILKERKLLRFVQ-TVPLELR 56
SP|P28549|CASA1_MACEU   MKLLIFSCLVTLALARPDALRLSIDRHFKHREL-----ENRLNEDPI-----PVSEASSS 50
SP|P04656|CASA1_CAVPO   MKLLILTCLVASAVAMPKFPFRHTELFQTQRGGSS---SSSSSEE------------RLK 45
SP|O62823|CASA1_BUBBU   MKLLILTCLVAVALARPKQPIKHQGLPQG-----------VLN-ENLLRFFVAPFPEVFG 48

SP|P47710-1|CASA1_HUMAN EEYMNGMNRQRNILREKQTDEIKDTRNEST-QNCVVAEPEKM----------ESSISSS- 90
SP|P47710-2|CASA1_HUMAN EEYMNGMNR-RNILREKQTDEIKDTRNEST-QNCVVAEPEKM----------ESSISSS- 89
SP|P47710-3|CASA1_HUMAN EEYMNGMNRQRNILREKQTDEIK---------NCVVAEPEKM----------ESSISSS- 82
SP|P47710-4|CASA1_HUMAN EEYMNGMNR-RNILREKQTDEIKDTRNEST-QNCVVAEPEKM----------ESSISSS- 89
SP|P02662|CASA1_BOVIN   KEKVNELSKD------IGSES--------T-EDQAMEDIKQM----------EAESISS- 82
SP|P19228|CASA1_MOUSE   QEFVNNMNRQRALLTEQ-NDEIKVTMDAAS-EEQAMASAQED----------SSIS-SSS 96
SP|P04653|CASA1_SHEEP   KENINELSKD------IGSES--------I-EDQAMEDAKQM----------KAGSSSS- 82
SP|P02661|CASA1_RAT     EEFVNNLNRQRELLTEQ-DNEIKITMDSSA-EEQATASAQED----------SSSSSSSS 99
SP|P18626|CASA1_CAPHI   KENINELSKD------IGSES--------T-EDQAMEDAKQM----------KAGSSSS- 82
SP|O97943|CASA1_CAMDR   QENIDELKDTRN-------EP--------T-EDHIMEDTERK----------E-SGSSS- 87
SP|P39035|CASA1_PIG     QEIINELN----------------------RNHGMEGHEQR----------G-SSSSS- 81
SP|P86272|CASA1_EQUAS   -EYINELNRQRELLKEKQKDE--------H-KEYLIEDPEQQ----------ESSSTSS- 77
SP|P09115|CASA1_RABIT   EEYVNELNRQRELLREKENEEIKGTRNEVT-EEHVLADRET-----------EASISSS- 103
SP|P28549|CASA1_MACEU   EESVHQLNRDRRPLEKYELDKYREDLKTSSSEEFVTPSTNERVRRQVEYNFNEEDSSAS- 109
SP|P04656|CASA1_CAVPO   EENIFKFDQQKEL-QRKQSEKIKEIIS---------ESTEQR----------EASSISS- 84
SP|O62823|CASA1_BUBBU   KEKVNELSTD------IGSES--------T-EDQAMEDIKQM----------EAESISS- 82
```

Figure S4. Protein sequence alignment for alpha-S1-casein across different species. The red colored letters indicate the novel *N*-glycosylation site found in human alpha-S1-casein. This site is not present in isoform 3 of human alpha-S1-casein. It seems to be unique for the human orthologue and is not present in related mammalian species.

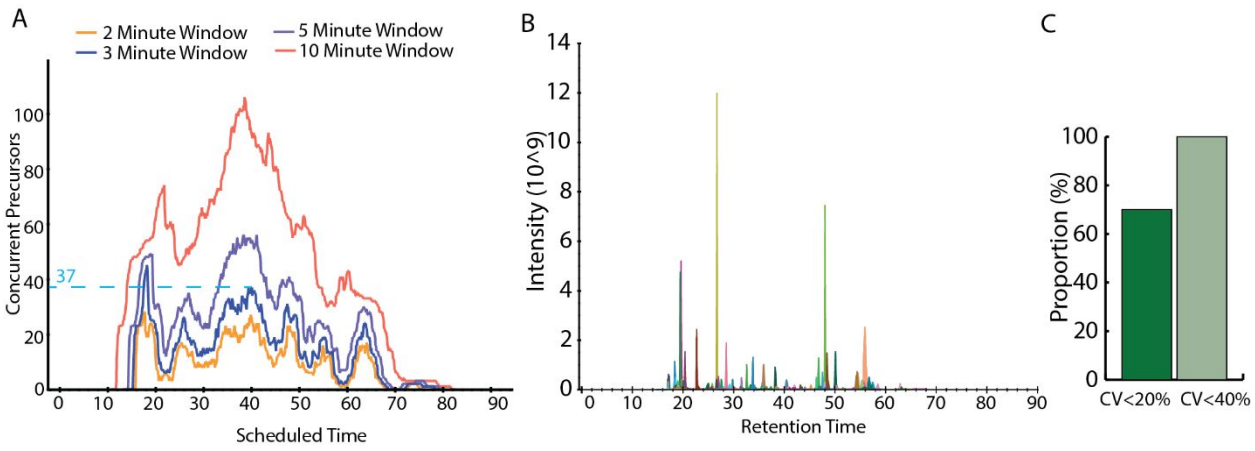

Figure S5. Experimental parameters used in the scheduled targeted assay for glycopeptide quantification. A. With the chosen 3 minutes window, the concurrent precursors were at a maximum 37 (blue line). Taking 54 ms as the ion injection time, a mass resolution setting of 30,000 and an average peak width of 30 s, at least 10 data points were taken over each chromatographic peak. B. Overall chromatogram of targeted glycopeptides. C. The CVs of all spiked PRTC peptides were calculated for all the runs, 70% of them had a CV less than 20%, another 30% had a CV less than than 40%. CV: coefficient of variation; PRTC: Pierce Peptide Retention Time Calibration Mixture.