Supplementary Information

Site-specific *N*-linked glycosylation analysis of human carcinoembryonic antigen by sheathless capillary electrophoresis - tandem mass spectrometry

Laura Pont[a,1], Valeriia Kuzyk[b,c,1], Fernando Benavente[a,*], Victoria Sanz-Nebot[a], Oleg A. Mayboroda[b], Manfred Wuhrer[b], Guinevere S.M. Lageveen-Kammeijer[b,*]

[a] Department of Chemical Engineering and Analytical Chemistry, University of Barcelona, Barcelona, Spain

[b] Center for Proteomics and Metabolomics, Leiden University Medical Center, Leiden, The Netherlands

[c] Division of Bioanalytical Chemistry, Vrije Universiteit Amsterdam, The Netherlands

[1] *Shared first author*

[*] *Corresponding authors: Dr. G.S.M. Lageveen-Kammeijer, g.s.m.kammeijer@lumc.nl, Tel: (+31) 715266993 and Dr. F. Benavente, fbenavente@ub.edu, Tel: (+34) 934039116*

Table of Contents

FIGURE S-1 – CE-MS/MS FRAGMENTATION SPECTRA

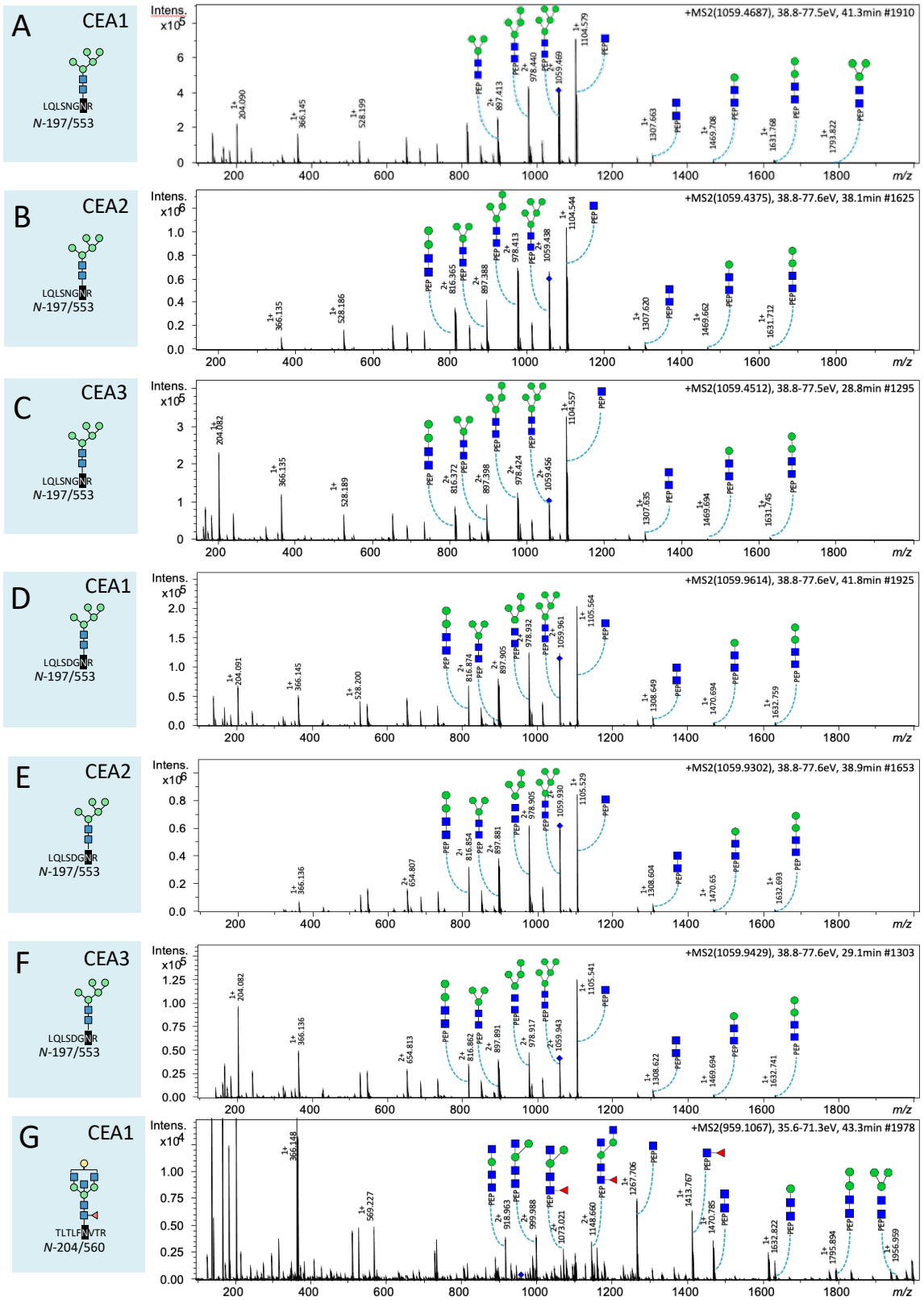

Figure S-1 CE-ESI-MS/MS fragmentation spectra of the most abundant *N*-glycopeptides per glycosylation site that were fragmented. The following precursor ions were fragmented **(A)** 1059.469 $^{2+}$ (H5N2), **(B)** 1059.438 $^{2+}$ (H5N2), **(C)** 1059.451 $^{2+}$ (H5N2), **(D)** 1059.961 $^{2+}$ (H5N2), **(E)** 1059.930 $^{2+}$ (H5N2), **(F)** 1059.943 $^{2+}$ (H5N2) and **(G)** 959.107 $^{3+}$ (H4N5F1). Blue diamond: precursor ion, blue square: *N*-acetylglucosamine, green circle: mannose, yellow circle: galactose, red triangle: fucose, right pointing pink diamond: α2,6-linked *N*-acetylneuraminic acid, left pointing pink diamond: α2,3-linked *N*-acetylneuraminic acid, PEP: peptide backbone as indicated in the blue box.

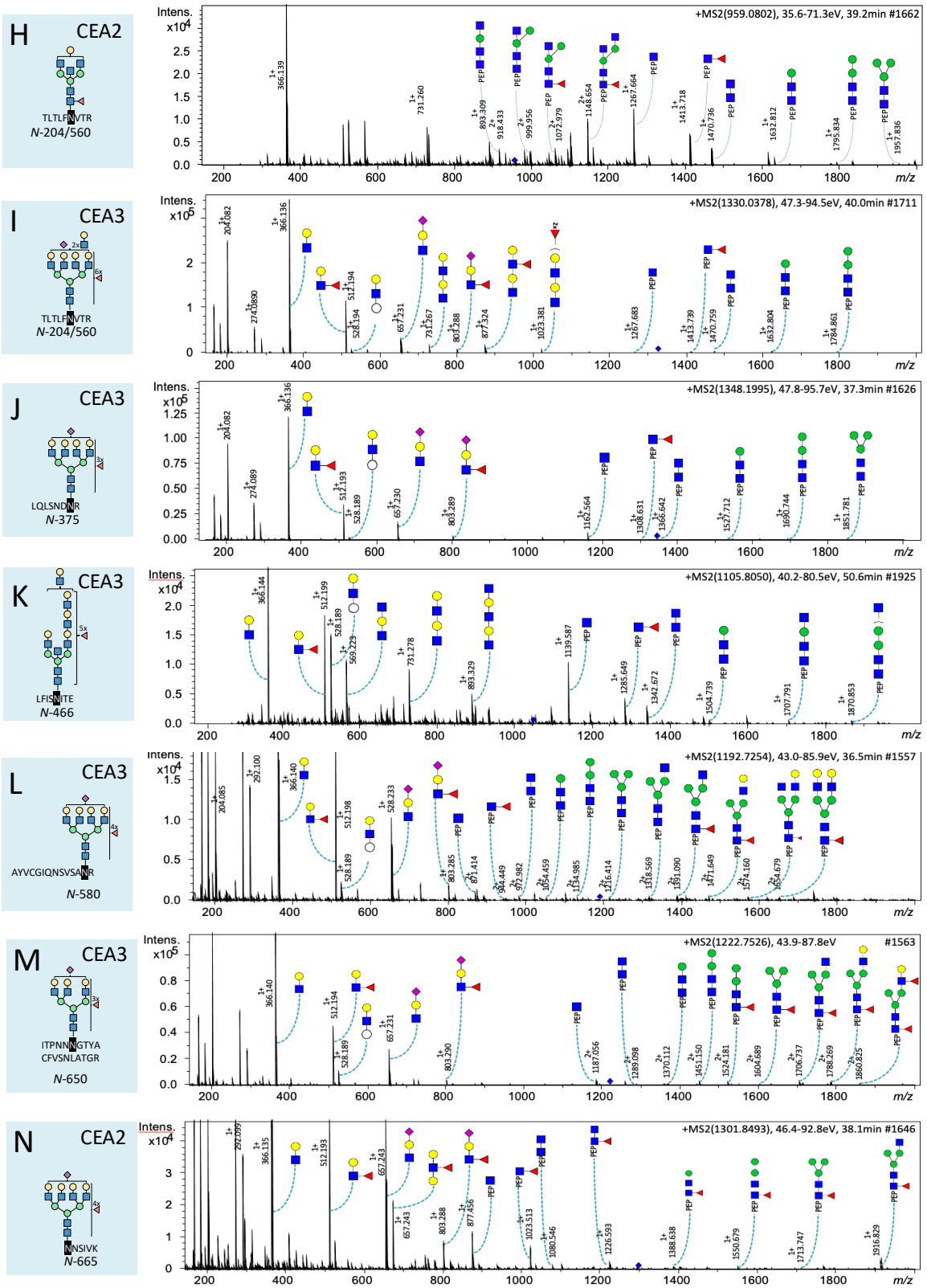

Figure S-1 (continued). CE-ESI-MS/MS fragmentation spectra of the most abundant *N*-glycopeptides per glycosylation site that were fragmented. The following precursor ions were **(H)** 959.080 $^{3+}$(H4N5F1), **(I)** 1330.038 $^{4+}$ (H9N8F6S1), **(J)** 1348.200 $^{3+}$ (H7N6F3S1), **(K)** 1105.805 $^{4+}$ (H8N8F5), **(L)** 1192.725 $^{4+}$ (H7N6F4S1), **(M)** 1222.753 $^{4+}$ (H6N5F3S1) and **(N)** 1301.849 $^{3+}$ (H7N6F4S1). Blue diamond: precursor ion, blue square: *N*-acetylglucosamine, green circle: mannose, yellow circle: galactose, red triangle: fucose, right pointing pink diamond: α2,6-linked *N*-acetylneuraminic acid, left pointing pink diamond: α2,3-linked *N*-acetylneuraminic acid, PEP: peptide backbone as indicated in the blue box.

FIGURE S-2 – DATA ASSESSMENT BY DATAANALYSIS AND SKYLINE

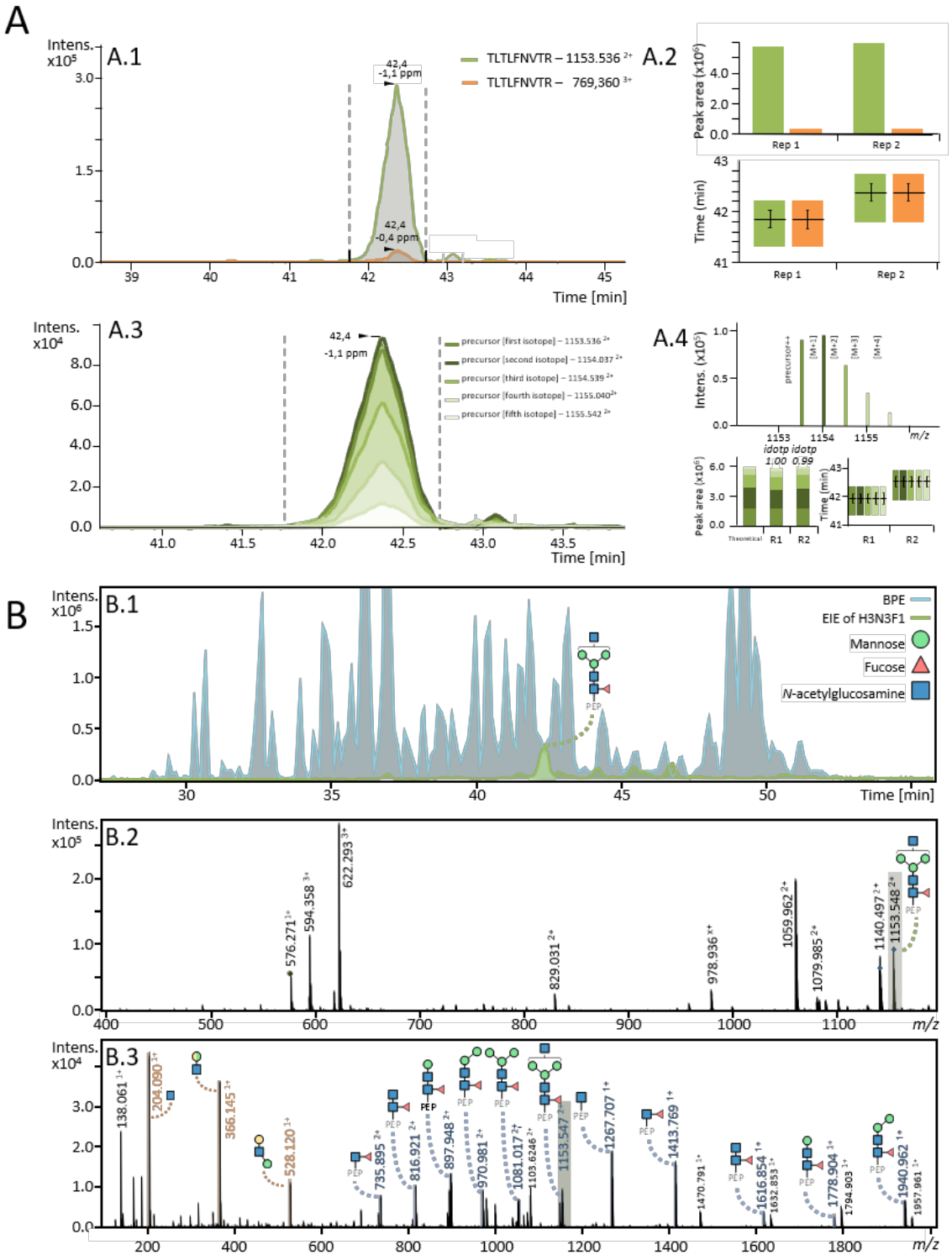

Figure S-2 Peak peaking, quantitation and annotation processes in Skyline (A) and Bruker DataAnalysis (B) softwares on the example of TLTLFNVTR H3N3F1 glycopeptide. (**A1**) EIE of the $[M+H]^{2+}$ and $[M+3H]^{3+}$ molecular ions (CEA1, Replicate 2). (**A2**) Quality metrics for both ions in all replicates, such as migration times and peak areas. (**A3**) EIEs of all the [M+H]2+ isotopes (**A4**) Quality metrics for all isotopes in all replicates, such as isotopic pattern, migration times and peak areas. (**B1**) EIE (green trace) of the theoretical calculated, taking into account the first isotope. The BPE is highlighted in blue. (**B2**) MS1 mass spectrum at 42.2 min. The $[M+H]^{2+}$ parent ion is illustrated with a green highlight. (**B3**) MS2 mass spectrum of $[M+H]^{2+}$ parent ion. Diagnostic oxonium ions (B-ions) are highlighted in orange and Y-ions are highlighted in blue. Green highlight: precursor ion and PEP: peptide (TLTLFNVTR).

FIGURE S-3 – PROTEIN COVERAGE

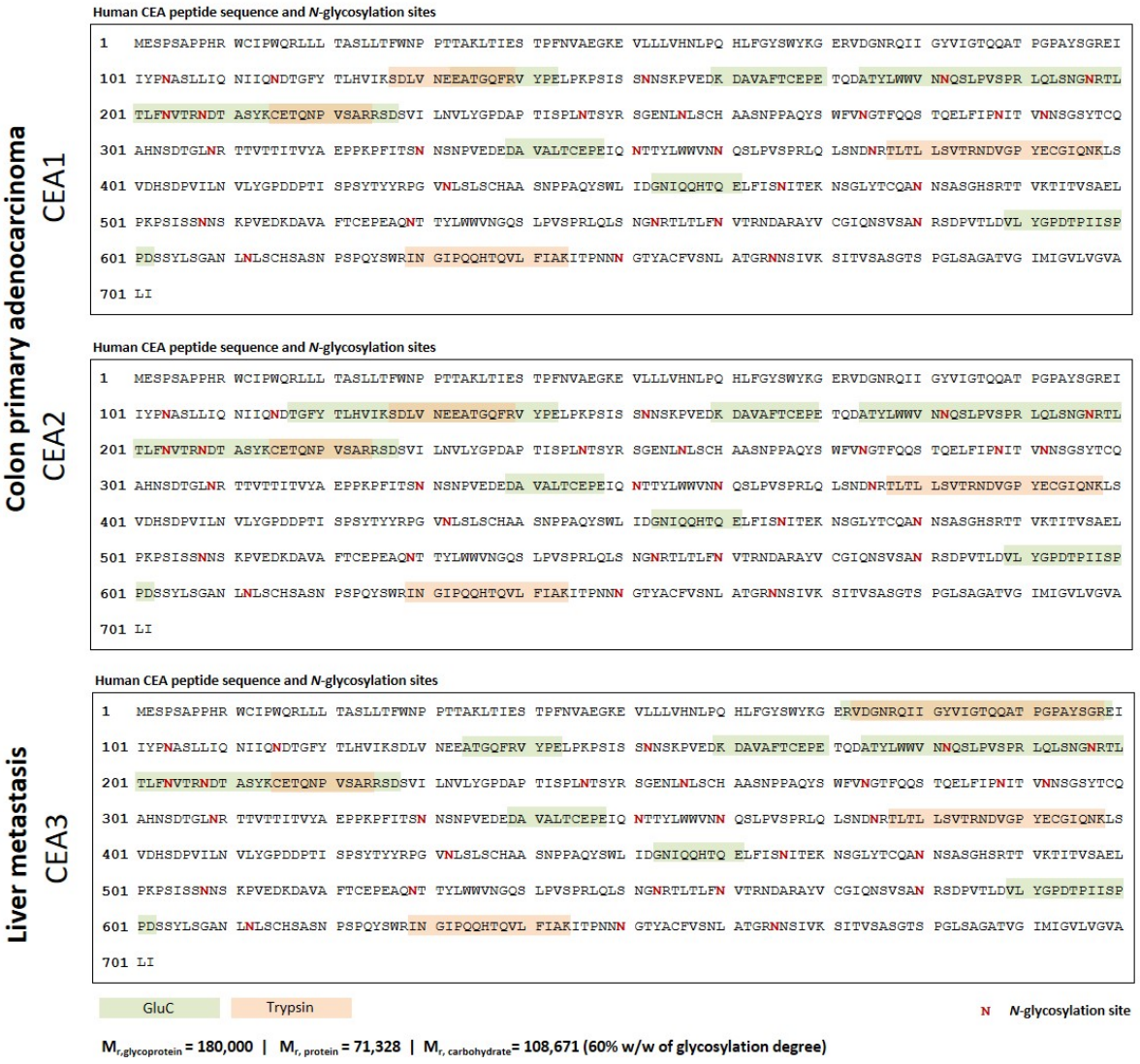

Figure S-3. Human carcinoembryonic antigen (CEACAM5_HUMAN) sequence. Putative *N*-glycosylation sites are shown in bold red. Peptides that were detected after specific proteases digestion are highlighted in green (GluC) and orange (trypsin). The total M_r of the glycosylated protein is reported to vary between 150,000 and 200,000, the mass calculations are based on the most prominent M_r (180,000).

FIGURE S-4 – DERIVED TRAITS PER GLYCOSYLATION SITE

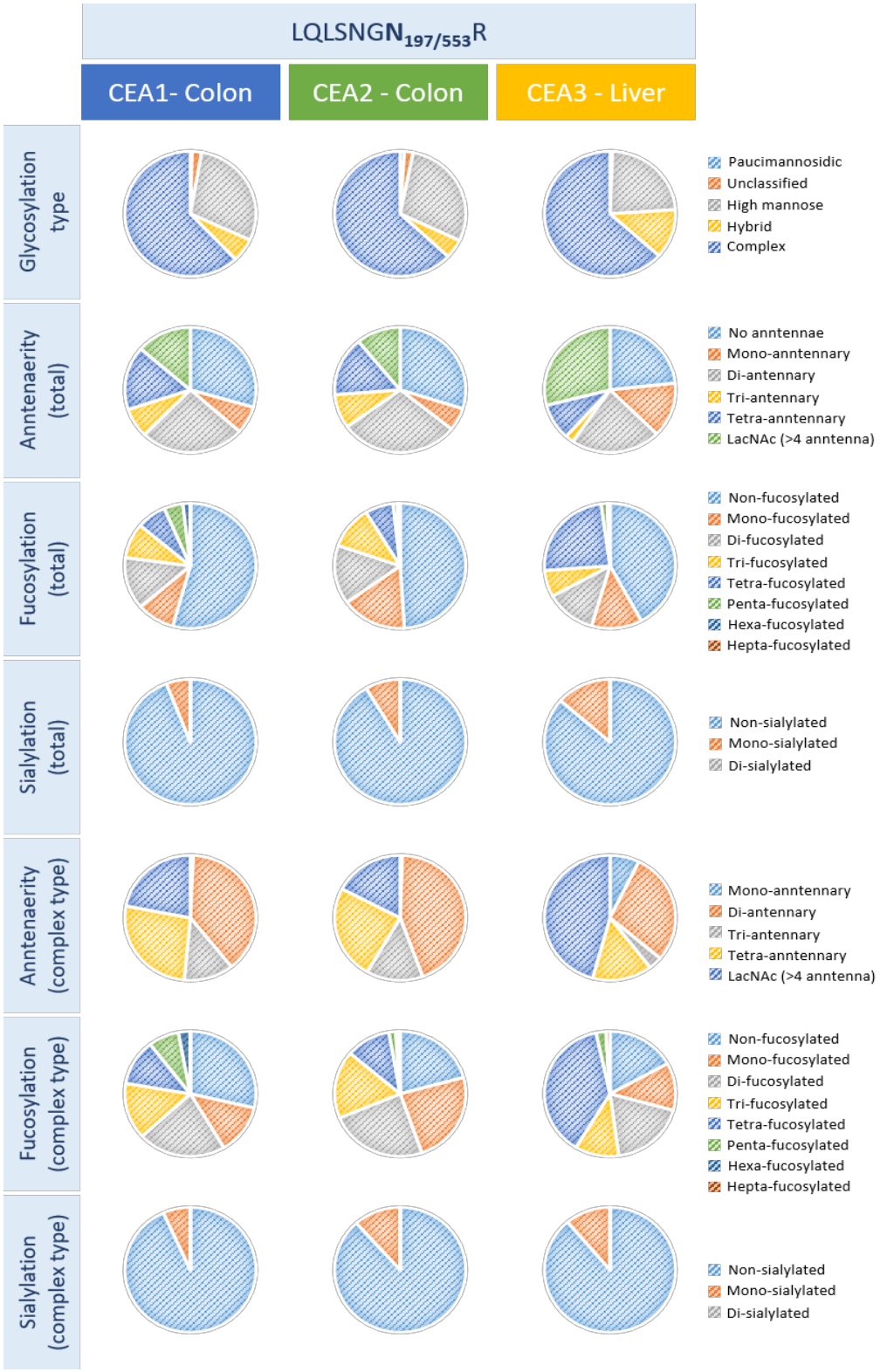

Figure S-4. Derived traits of glycosylation site $N_{197/553}$ illustrated for all the three CEA samples after digestion with trypsin and analysis with CE-MS/MS. A "trait" is the relative abundance of all glycopeptide glycoforms observed within the sample complying with specific characteristics (e.g. Sialylation: non-sialylated, mono-sialylated or di-sialylated). The different glycosylation types were investigated as well as antennarity, fucosylation and sialylation present on all glycans (total) and within the complex types (complex). All identified glycan species and the derived traits calculations can be found in **Supplementary Information, Table S-2 and Table S-4**, respectively.

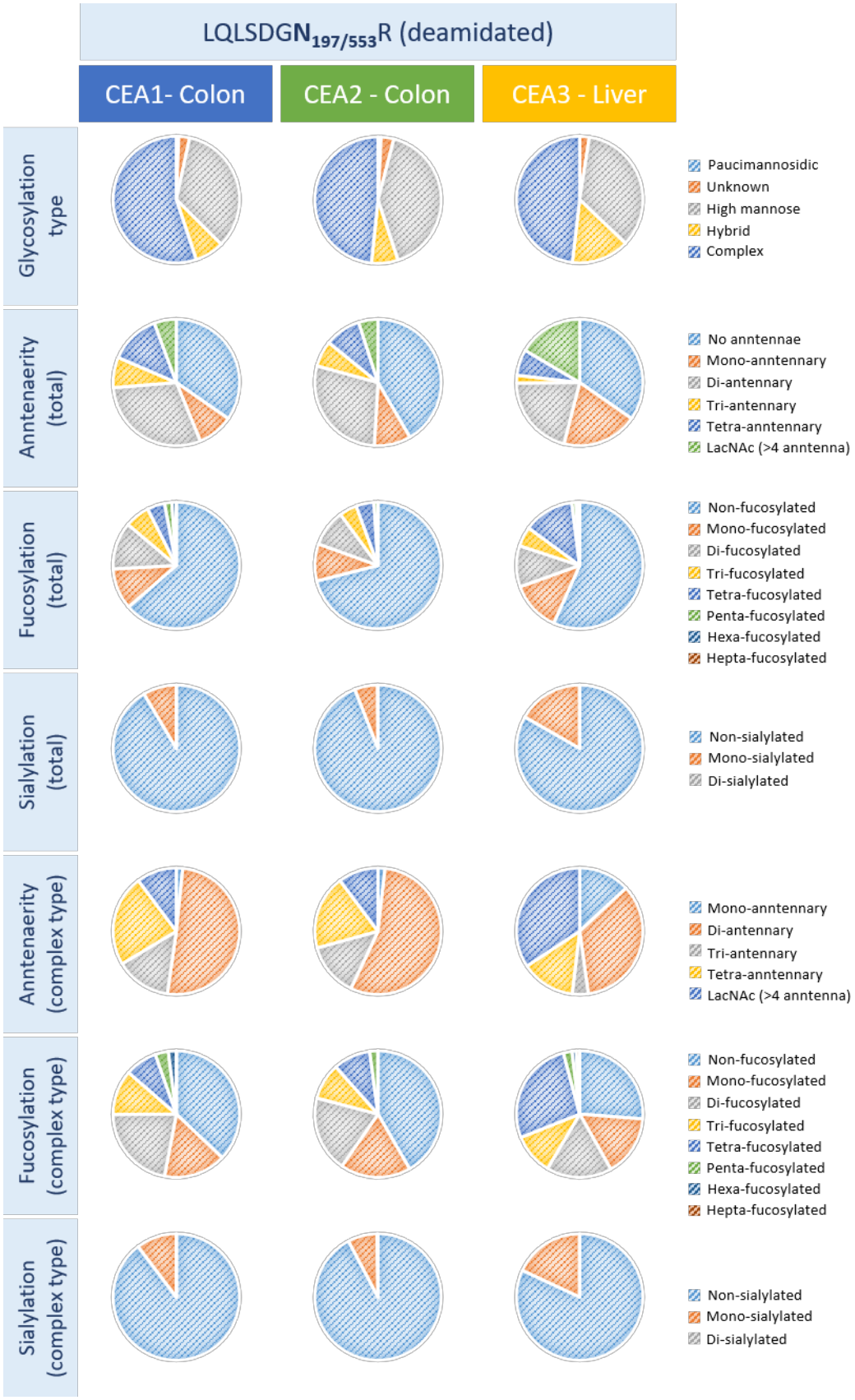

Figure S-4 (continued). Derived traits of glycosylation site $N_{197/553}$ (deamidated) illustrated for all the three CEA samples after digestion with trypsin and analysis with CE-MS/MS. *A* "trait" is the relative abundance of all glycopeptide glycoforms observed within the sample complying with specific characteristics (e.g. Sialylation: non-sialylated, mono-sialylated or di-sialylated). The different glycosylation types were investigated as well as antennarity, fucosylation and sialylation present on all glycans (total) and within the complex types (complex). All identified glycan species and the derived traits calculations can be found in **Supplementary Information, Table S-2 and Table S-4**, respectively.

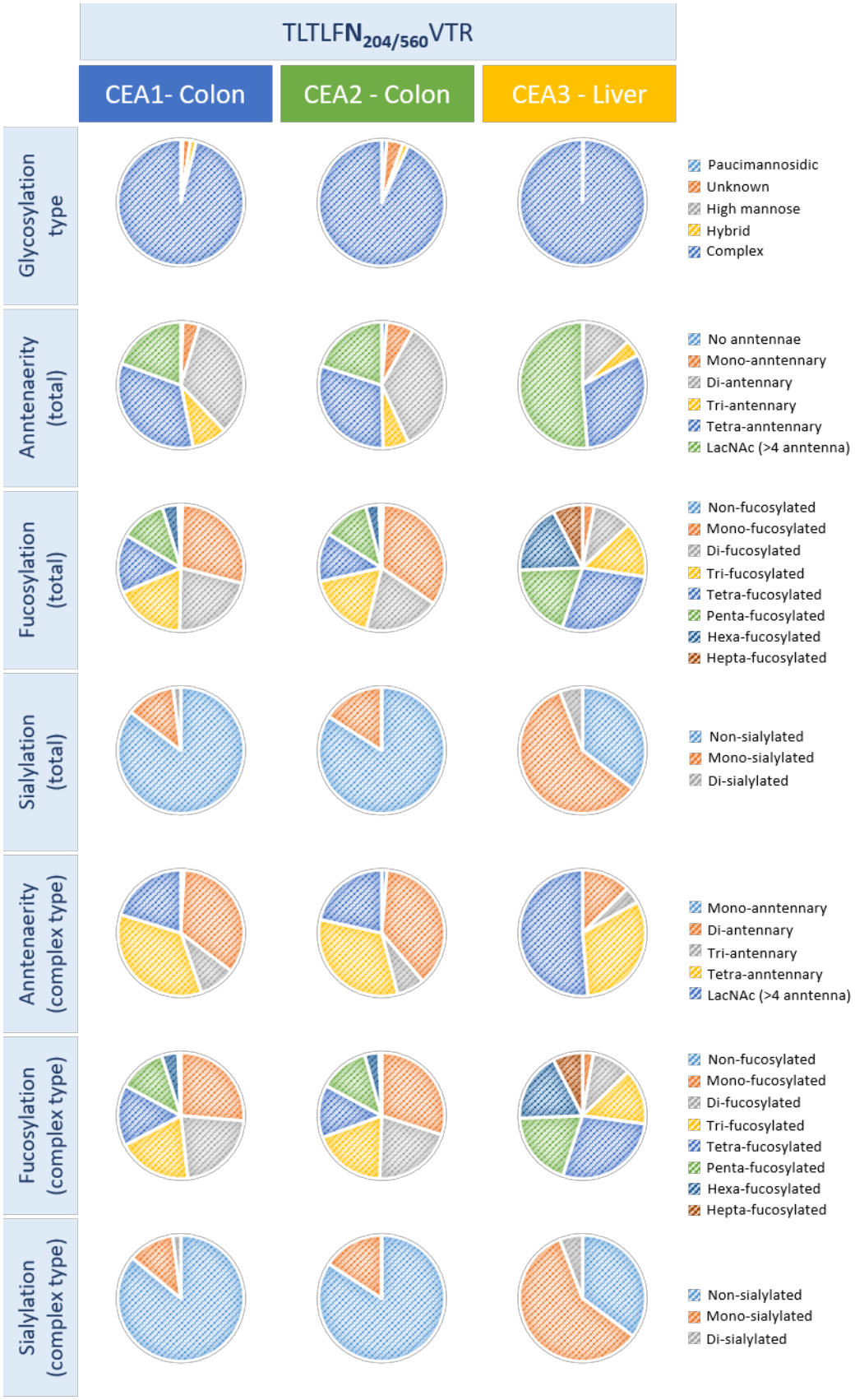

Figure S-4 (continued). Derived traits of glycosylation site $N_{204/560}$ illustrated for all the three CEA samples after digestion with trypsin and analysis with CE-MS/MS. *A* "trait" is the relative abundance of all glycopeptide glycoforms observed within the sample complying with specific characteristics (e.g. Sialylation: non-sialylated, mono-sialylated or di-sialylated). The different glycosylation types were investigated as well as antennarity, fucosylation and sialylation present on all glycans (total) and within the complex types (complex). All identified glycan species and the derived traits calculations can be found in **Supplementary Information, Table S-2 and Table S-4**, respectively.

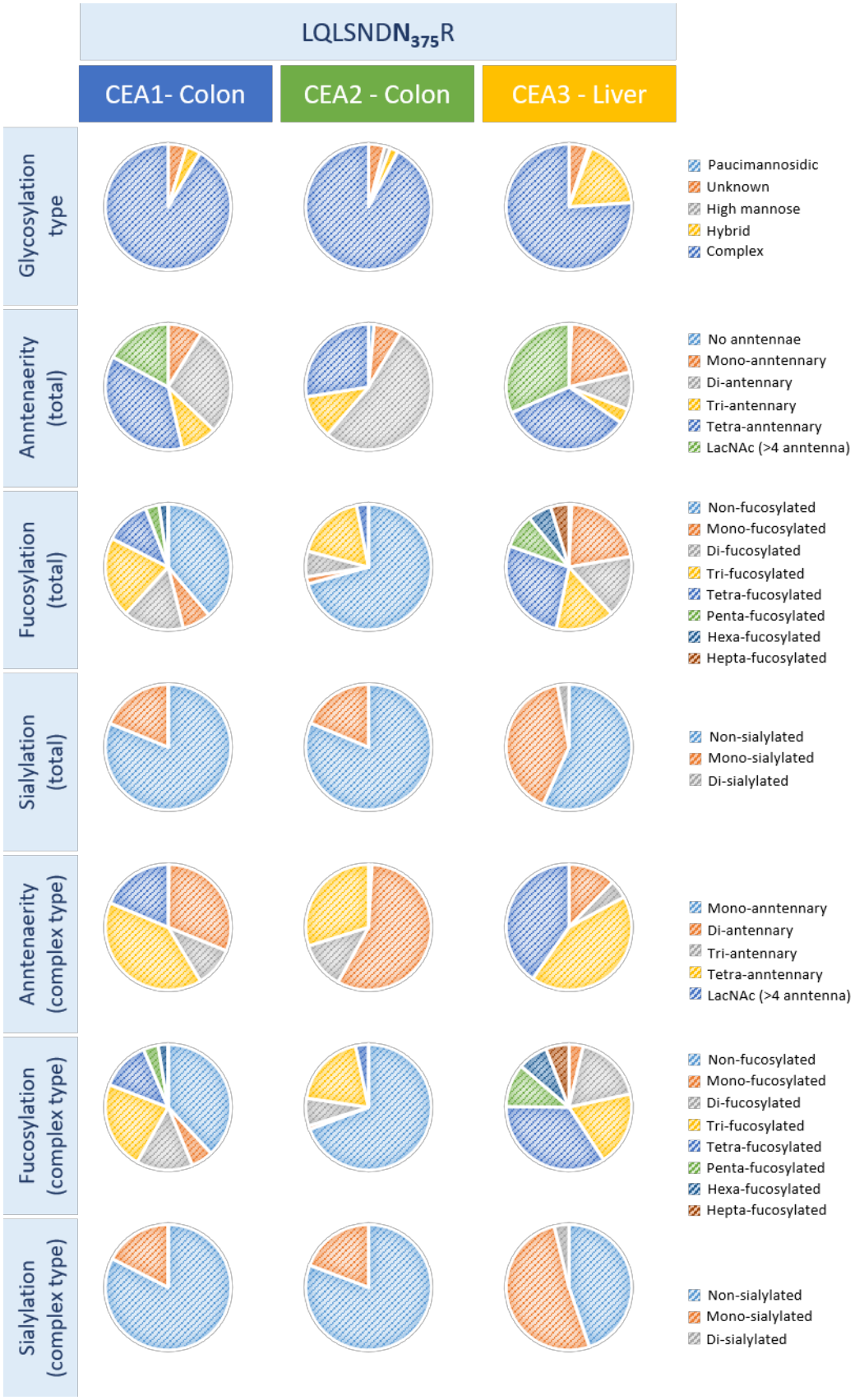

Figure S-4 (continued). Derived traits of glycosylation site N_{375} illustrated for all the three CEA samples after digestion with trypsin and analysis with CE-MS/MS. *A* "trait" is the relative abundance of all glycopeptide glycoforms observed within the sample complying with specific characteristics (e.g. Sialylation: non-sialylated, mono-sialylated or di-sialylated). The different glycosylation types were investigated as well as antennarity, fucosylation and sialylation present on all glycans (total) and within the complex types (complex). All identified glycan species and the derived traits calculations can be found in **Supplementary Information, Table S-2 and Table S-4**, respectively.

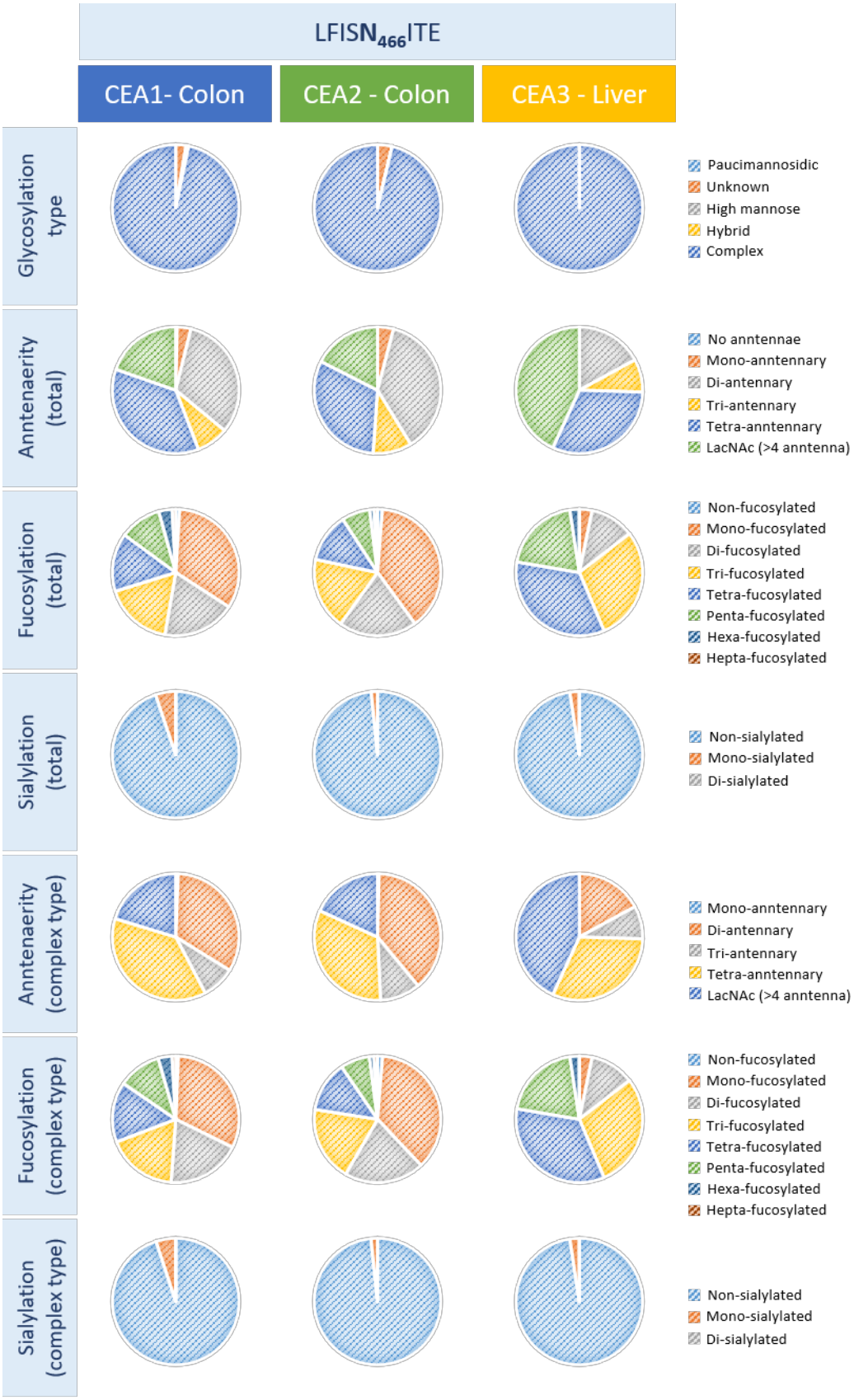

Figure S-4 (continued). Derived traits of glycosylation site N_{466} illustrated for all the three CEA samples after enzymatic digestion with GluC and analysis with CE-MS/MS. *A* "trait" is the relative abundance of all glycopeptide glycoforms observed within the sample complying with specific characteristics (e.g. Sialylation: non-sialylated, mono-sialylated or di-sialylated). The different glycosylation types were investigated as well as antennarity, fucosylation and sialylation present on all glycans (total) and within the complex types (complex). All identified glycan species and the derived traits calculations can be found in **Supplementary Information, Table S-2 and Table S-4**, respectively.

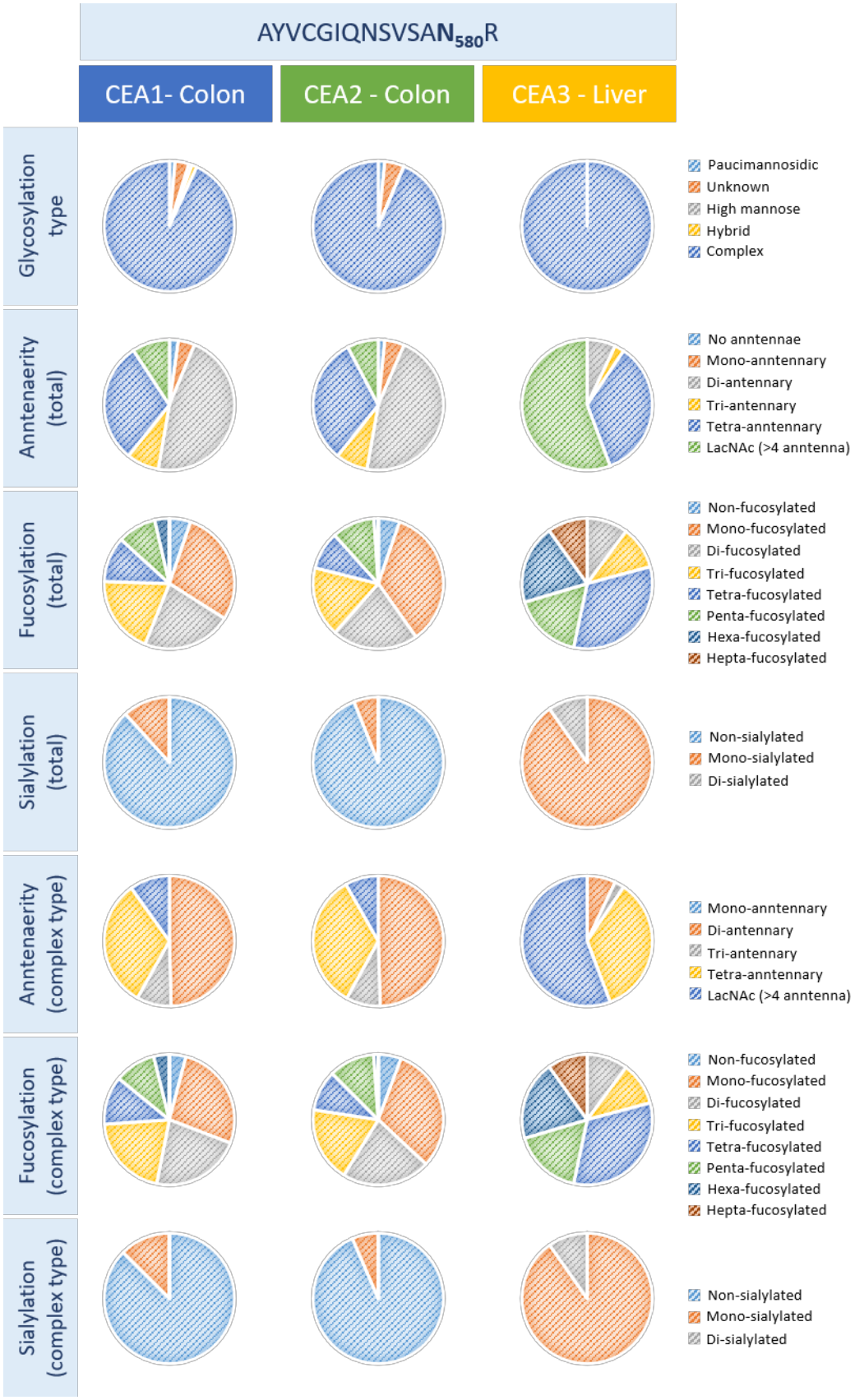

Figure S-4 (continued). Derived traits of glycosylation site N_{580} illustrated for all the three CEA samples after enzymatic digestion with trypsin and analysis with CE-MS/MS. *A* "trait" is the relative abundance of all glycopeptide glycoforms observed within the sample complying with specific characteristics (e.g. Sialylation: non-sialylated, mono-sialylated or di-sialylated). The different glycosylation types were investigated as well as antennarity, fucosylation and sialylation present on all glycans (total) and within the complex types (complex). All identified glycan species and the derived traits calculations can be found in **Supplementary Information, Table S-2 and Table S-4**, respectively.

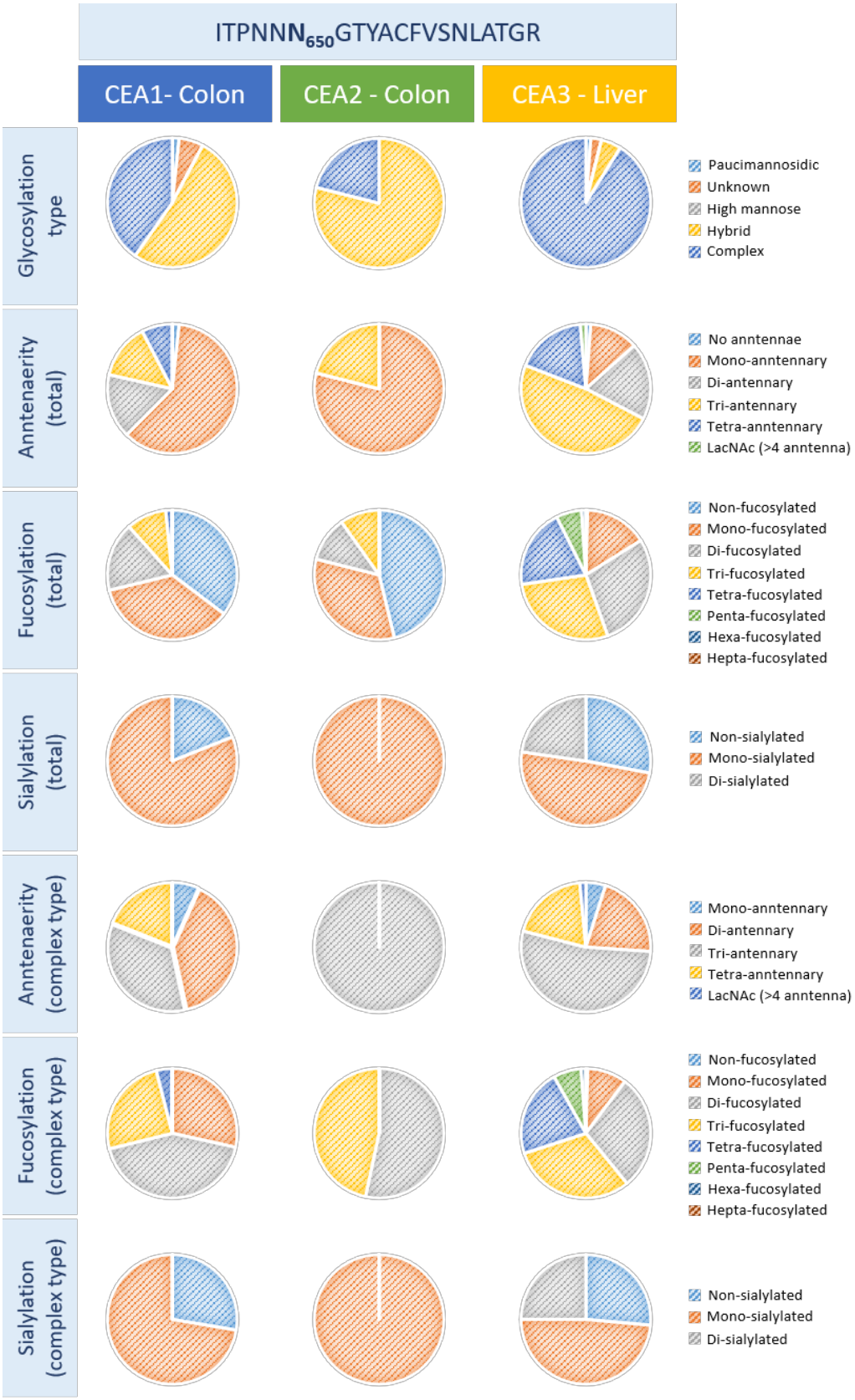

Figure S-4 (continued). Derived traits of glycosylation site N_{650} illustrated for all the three CEA samples after enzymatic digestion with trypsin and analysis with CE-MS/MS. *A* "trait" is the relative abundance of all glycopeptide glycoforms observed within the sample complying with specific characteristics (e.g. Sialylation: non-sialylated, mono-sialylated or di-sialylated). The different glycosylation types were investigated as well as antennarity, fucosylation and sialylation present on all glycans (total) and within the complex types (complex). All identified glycan species and the derived traits calculations can be found in **Supplementary Information, Table S-2 and Table S-4**, respectively.

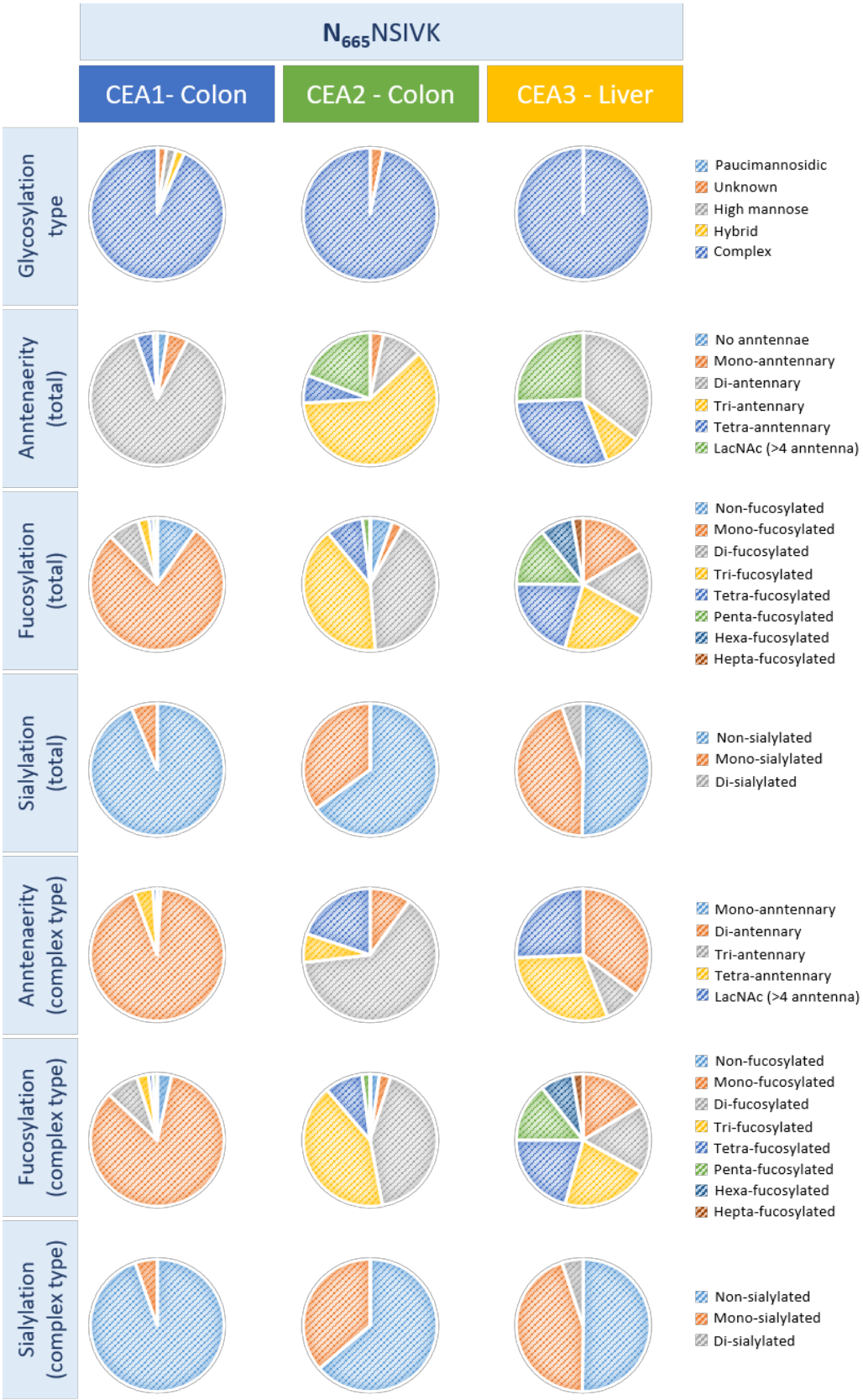

Figure S-4 (continued). Derived traits of glycosylation site N_{665} illustrated for all the three CEA samples after enzymatic digestion with trypsin and analysis with CE-MS/MS. *A* "trait" is the relative abundance of all glycopeptide glycoforms observed within the sample complying with specific characteristics (e.g. Sialylation: non-sialylated, mono-sialylated or di-sialylated). The different glycosylation types were investigated as well as antennarity, fucosylation and sialylation present on all glycans (total) and within the complex types (complex). All identified glycan species and the derived traits calculations can be found in **Supplementary Information, Table S-2 and Table S-4**, respectively.

FIGURE S-5 – RELATIVE ABUNDANCE OF THE TOP 10 GLYCOFORMS PER GLYCOSYLATION SITE

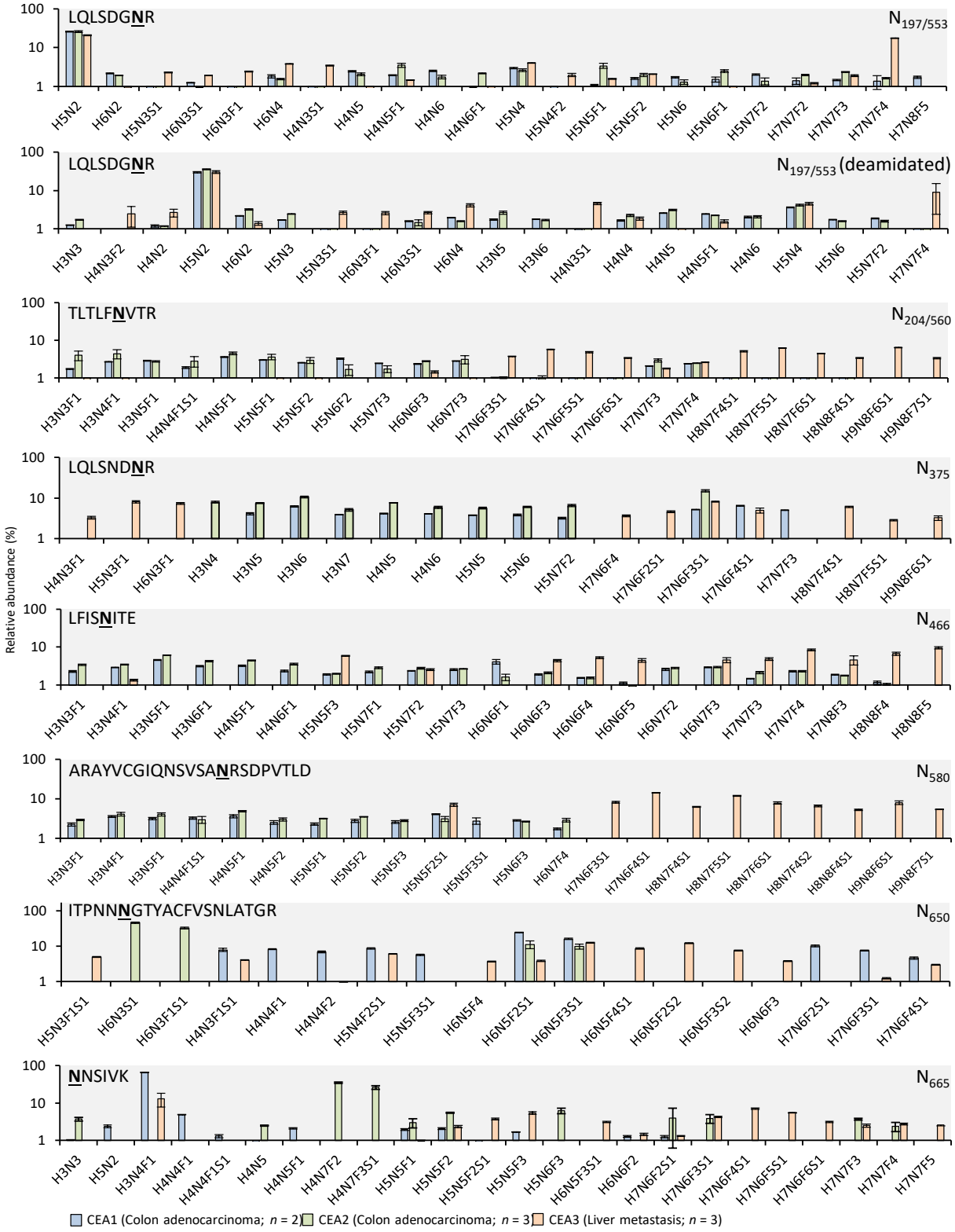

Figure S-5. Relative abundances of the 10 most abundant glycoforms of the quantitatively characterized *N*-glycosylation sites (CEA1 is plotted in blue, CEA2 in green and CEA3 in orange). Peak area values were normalized by *N*-glycosylation site for all samples (as percentage of all glycoforms peak areas detected per site). Y-axis is presented in logarithmic scale and error bars indicate the standard deviation (n=3). H: hexose, N: *N*-acetylhexosamine, F: fucose and S: *N*-acetyl neuraminic acid.