Abstract

Motivation: Glycosylation is one of the most heterogenous and complex post-translational modifications, but.
Results: These are the resutls for this article.

Application of Network Smoothing to Glycan LC-MS Profiling

Joshua Klein

July 9, 2017

1 Introduction

Glycosylation is one of the most pervasive forms of post-translational modification.

4 Discussion

4.1 Glycan Assignment Performance

4.1.1 Simple Samples

In *20151002-02-IGG* and *20150930-06-AGP* the known common glycoforms were assigned without ambiguity, both with and without formate adducts, confirmed manually. In these cases, network smoothing is not necessary. The estimates of τ in *20151002-02-IGG* are low because the number of glycan compositions observed is low and the overlap in $\mathbf{A}_o$ is large. In *20150930-06-AGP* , there are many more compositions to assign, which in turn leads to larger τ estimates, with good support for all observations with the exception of {`Hex:5; HexNAc:4; Neu5Ac:1`} which depends upon the asialo-bi-antennary neighborhoods, and is also the only observed member of this neighborhoods.

4.1.2 Influenza Strains

In *20141031-07-Phil-82* , the unregularized case contains some ambiguous matches where the biological context implies they should not be possible, {`Hex:7; HexNAc:6; Neu5Ac:4`}, or where the biological context and neighboring observations support the presence of a glycan composition but the evidence does not satisfy the scoring function, {`Hex:10; HexNAc:2`}, {`Hex:6; HexNAc:5`}, and {`Fuc:1; Hex:7; HexNAc:6`}. By applying the smoothing procedure with the parameters automatically estimated with grid search (Table 6), {`Hex:7; HexNAc:6; Neu5Ac:4`} was eliminated, while {`Hex:10; HexNAc:2`} and {`Hex:6; HexNAc:5`} were boosted into a higher confidence score range. The change to {`Fuc:1; Hex:7; HexNAc:6`} was insufficient to reach a high confidence score range, and the score for {`Fuc:1; Hex:8; HexNAc:7`} was dropped from a plausible score range to a low confidence range. Both these larger asialo-N-glycans have been previously assigned in Khatri *et al.* (2016a), but the automated procedure drops these compositions while estimating $\hat{\gamma}$ leading to empty neighborhoodss when estimating τ. A user-generated τ would contain a value greater than 5 in those neighborhoods.

This would be another table entry in tbl:phil82_score_table, already computed but uncertain how to integrate into the flow yet

We are justified in {`Hex:7; HexNAc:6; Neu5Ac:4`}'s removal based on its lack of supporting intermediary glycoforms and relatives above the selected $\hat{\gamma}$. If its own score exceeded $\hat{\gamma}$, it would itself result in a non-zero value for its related neighborhoods, providing itself with a non-zero minimum value and adjusting its rate of decay with λ. This is not to say that the LC-MS evidence that was observed is not real signal, merely that the assignment of that signal the glycan composition {`Hex:7; HexNAc:6; Neu5Ac:4`} is unlikely given the context.

The related *20141128-11-Phil-82* sample we see a similar pattern of glycoforms though with a wider range of fucosylation. In this case, we also have chemical noise from permethylation and a low degree of ammonium adduction, which can lead to more spurious matches. We match several multiply sialylated glycan compositions in this sample which do not satisfy any of the neighborhood conditions in Table 4, which results in their score decaying rapidly as $\lambda \to 1$. If these compositions were viable under the user's glycome model, then a different set of neighborhood rules would need to be specified which covered this group. In this case, several of them are ambiguous assignments of the same signal due to the mass shift imposed by ammonium (17.026 Da, `H3 N1`) compared to a proton adduct, is nearly the same as the difference between a permethylated **Neu5Ac** and permethylated **FucHex** (17.015 Da, `H3 C1 O1 N-1`). In other cases, monosialylated compositions are either not eliminated or receive a larger score after smoothing because they are connected to the asialo neighborhoods, which are strongly supported in this glycome. We manually confirmed that the signal assigned to {`Fuc:1; Hex:5; HexNAc:4; Neu5NAc:1`} is not a deconvolution artefact, but the signal for {`Fuc:2; Hex:5; HexNAc:4; Neu5NAc:1`} is partially overlapped and difficult to manually separate. While this connection between monosialo and asialo forms may make sense in some cases, the link may not be appropriate for the biological context of this sample where a viral protein Neuraminadase removes **Neu5Ac**, and the user could redefine the neighborhoods to omit that overlap. Similar commentary can be made for *20141101-04-Phil-BS* , which while not permethylated and ammoniated, shows the presence of sialylated compositions. In

4.2 Utility of Network Smoothing

References

Akune, Y., Lin, C.-H., Abrahams, J. L., Zhang, J., Packer, N. H., Aoki-Kinoshita, K. F., and Campbell, M. P. (2016). Comprehensive analysis of the N-glycan biosynthetic pathway using bioinformatics to generate UniCorn: A theoretical N-glycan structure database. *Carbohydrate Research*, **431**, 56–63.

Belkin, M., Niyogi, P., and Sindhwani, V. (2006). Manifold regularization: A geometric framework for learning from labeled and unlabeled examples. *Journal of Machine Learning Research*, **7**(2006), 2399–2434.

Goldberg, D., Bern, M., North, S. J., Haslam, S. M., and Dell, A. (2009). Glycan family analysis for deducing N-glycan topology from single MS. *Bioinformatics*, **25**(3), 365–371.

Kaur, P. and O'Connor, P. B. (2006). Algorithms for automatic interpretation of high resolution mass spectra. *Journal of the American Society for Mass Spectrometry*, **17**(3), 459–468.

Kessner, D., Chambers, M., Burke, R., Agus, D., and Mallick, P. (2008). ProteoWizard: Open source software for rapid proteomics tools development. *Bioinformatics*, **24**(21), 2534–2536.

Khatri, K., Klein, J. A., White, M. R., Grant, O. C., Lemarie, N., Woods, R. J., Hartshorn, K. L., Zaia, J., Leymarie, N., Woods, R. J., Hartshorn, K. L., and Zaia, J. (2016a). Integrated omics and computational glycobiology reveal structural basis for Influenza A virus glycan microheterogeneity and host interactions. *Molecular & cellular proteomics : MCP*, **13975**(615).

Khatri, K., Klein, J. A., and Zaia, J. (2016b). Use of an informed search space maximizes confidence of site-specific assignment of glycoprotein glycosylation. *Analytical and Bioanalytical Chemistry*.

Kronewitter, S. R., Slysz, G. W., Marginean, I., Hagler, C. D., LaMarche, B. L., Zhao, R., Harris, M. Y., Monroe, M. E., Polyukh, C. A., Crowell, K. L., Fillmore, T. L., Carlson, T. S., Camp, D. G., Moore, R. J., Payne, S. H., Anderson, G. a., and Smith, R. D. (2014). GlyQ-IQ: Glycomics quintavariate-informed quantification with high-performance computing and glycogrid 4D visualization. *Analytical Chemistry*, **86**(13), 6268–6276.

Liu, X., Inbar, Y., Dorrestein, P. C., Wynne, C., Edwards, N., Souda, P., Whitelegge, J. P., Bafna, V., and Pevzner, P. A. (2010). Deconvolution and database search of complex tandem mass spectra of intact proteins: a combinatorial approach. *Molecular & cellular proteomics : MCP*, **9**(12), 2772–2782.

Martens, L., Chambers, M., Sturm, M., Kessner, D., Levander, F., Shofstahl, J., Tang, W. H., Römpp, A., Neumann, S., Pizarro, A. D., Montecchi-Palazzi, L., Tasman, N., Coleman, M., Reisinger, F., Souda, P., Hermjakob, H., Binz, P.-a., and Deutsch, E. W. (2011). mzML–a community standard for mass spectrometry data. *Molecular & cellular proteomics : MCP*, **10**(1), R110.000133.

Maxwell, E., Tan, Y., Tan, Y., Hu, H., Benson, G., Aizikov, K., Conley, S., Staples, G. O., Slysz, G. W., Smith, R. D., and Zaia, J. (2012). GlycReSoft: a software package for automated recognition of glycans from LC/MS data. *PloS one*, **7**(9), e45474.

Senko, M. W., Beu, S. C., and McLafferty, F. W. (1995). Determination of monoisotopic masses and ion populations for large biomolecules from resolved isotopic distributions. *Journal of the American Society for Mass Spectrometry*, **6**(4), 229–233.

Stanley, P., Schachter, H., and Taniguchi, N. (2009). *N-Glycans*. Cold Spring Harbor Laboratory Press.

Varki, A. and Schauer, R. (2009). *Sialic Acids*. Cold Spring Harbor Laboratory Press.

Yu, C.-Y. C.-Y., Mayampurath, A., Hu, Y., Zhou, S., Mechref, Y., and Tang, H. (2013). Automated annotation and quantification of glycans using liquid chromatography-mass spectrometry. *Bioinformatics*, **29**(13), 1706–1707.

Yu, T. and Peng, H. (2010). Quantification and deconvolution of asymmetric LC-MS peaks using the bi-Gaussian mixture model and statistical model selection. *BMC bioinformatics*, **11**(1), 559.