A Summary and Project Proposal **Matthew Wing**
Genomics & Bioinformatics Final Project

Glycomics: an integrated systems approach to structure-function relationships of glycans is a review paper that provides a blueprint for researchers interested in pursuing further research in the field of glycomics. The field of glycomics is an emerging body of research that has, according to the paper, garnered increasing interest due to the emerging evidence protein post-translational modifications such as glycosylation are having on cell phenotypes. The authors highlight the forces limiting the rapid advancement of research in this discipline and offer suggestions on how to effectively approach glycomics research.

As the title suggests, the key to fruitful glycan investigation lies in using a systems approach. This necessitates the need for researchers to examine glycans at both a whole-organism phenotype level and a more specific scope relating to glycan structure, biochemistry, and biosynthesis. In order for this combined approach to be successful, the authors emphasize that the many datasets being generated across the globe must be integrated. Thus, the authors review the current technologies of glycomics and discuss the datasets that each technology generates. They then look at how these datasets are interrelated and what can be gained through integration using a bioinformatics platform.

The purpose of employing a functional genetics approach to glycomics is to understand the biological role of glycans and their relationship to cellular and whole-organism phenotype. According to the paper, studies using this approach have utilized transgenetic technologies in order to understand the phenotypic effects of knocking out particular glycan biosynthesis enzymes. This approach has generated large quantities of new data, however, and the authors note that much more still needs to be accomplished in the way of annotating enzyme reaction

specificity. Therefore, to acquire a better grasp of biosynthetic enzyme function, researchers must “couple” the functional approach with observations from technologies designed to further probe glycan gene expression, structure, and biochemistry.

The authors highlight the importance of microarray technologies with respect to the advancement of genomics and are optimistic about the usefulness of this technology with respect to the study of glycan biosynthesis gene expression. They do, however, identify the unique challenges presented when studying the glycan-protein interactions. These hurdles have necessitated a very specific approach towards creating glycol-gene-based microarrays. The chemical complexity of glycans has delayed the effective characterization glycan structures. The paper briefly describes tools used for high-throughput analysis such as mass spec methods and also overviews the complementary technologies for more detailed structure characterization such as FT-ICR-MS. The overview the authors provide on the biochemical analysis of glycan-protein interactions center on the analysis of the glycan-glycan binding protein (GBP) interactions. The authors address the issues researchers face due to the unique characteristic of multivalency that glycans exhibit. They note that multiple new arrays are being developed in order to take the unique aspects of glycans into account.

The authors propose that the best way to integrate all the datasets and information on glycan function and structure is to develop a bioinformatics platform specific to glycomics. Although the data and information being pooled and organized is different from that of genomics, the basic idea of disseminating salient information across multiple datasets remains constant. The basis of using this bioinformatics model was taken from analogous studies done with respect to the Gene Ontology Consortium. Similar to other developmental aspects of the

glycan technology, creating a bioinformatics platform specific to glycomics can be modifications of similar through processes taken from the field of genomics and proteomics.

By developing an integrated bioinformatics platform, the authors stress that researchers will be able to ask more complicated questions that involve interdependence of datasets. Simplification of the user interfaces through user-friendly software, creative efforts to link orthogonal data (i.e. molecule page interface), and computational tools that support data mining efforts are all examples the authors provide as central aspects of the platform that need attention. The paper concludes by underlining broad challenges this field needs to face in order to advance glycomics research. These obstacles center on the advanced understanding of glycan diversity and glycan-protein interactions, which the authors emphasize can begin to be solved by utilizing and developing the tools mentioned in the review.

Overall, I believe the authors did an excellent job with explaining the purpose of each technology and why the data obtained from its usage would yield important information. I also think that the overview of how to create a bioinformatics platform near the end of the review effectively tied the paper together. It emphasized the point that the value of the datasets these technologies can provide will be most effectively tapped if everything is integrated using a bioinformatics platform. However, this bioinformatics section is also one area that could use more discussion with respect to clarifying the importance of studying object “relationships” and how exactly they “facilitate” integration. Although the authors mention that one must look at the relationship between primary objects and then the corresponding secondary objects which are borne out of the technologies used to study them, the wording of the relationship framework between these objects gets somewhat confusing. I think that the authors tried to condense too large and important of a concept within a small paragraph.

Another area that I would have liked to see the authors touch upon would have been their discussion on the parallels and differences relating glycomics to the study of genomics and proteomics. For example, it would have been helpful if they mentioned the applicability of automated algorithms used in genomics or proteomics to the field of glycomics von der Lieth, 175). Although there were descriptions of how certain technologies had to be tweaked for glycan-specific research, there was not as much big-picture discussion about the parallels in the approaches. I also think it would have been extremely informative if the authors had highlighted the problems and challenges that existed when genomics was first emerging as a field. This discussion, however, may have taken away from the main point of the review paper. With this in mind, perhaps the authors could revisit this question in another review paper.

Despite this being a review paper, I think it is extremely relevant and important to anyone interested not only in glycomics but also in genomics and cell/molecular biology. I would refer this paper to a broad audience, ranging from an undergraduate thinking about graduate school to a veteran researcher in genomics or cell/molecular biology that is looking to conduct research in an emerging field. This paper's description of glycomics definitely underscores the vast opportunity to make ground-breaking research and makes the idea of studying in this field very exciting. Because of the novelty of this field, a young graduate student may find the study of a particular technology such as glyco microarrays fascinating or, instead, may be attracted to this field as a result of the opportunity to identify novel ways to integrate all the new information. Likewise, a curious professor might want to apply his/her knowledge of say, genomics, to the development of parallel technologies related to glycans. Perhaps a cell biologist would like to take a more system-oriented approach to understanding phenotype; the field of glycomics would provide him/her with this opportunity. If this paper were not enough impetus for someone to

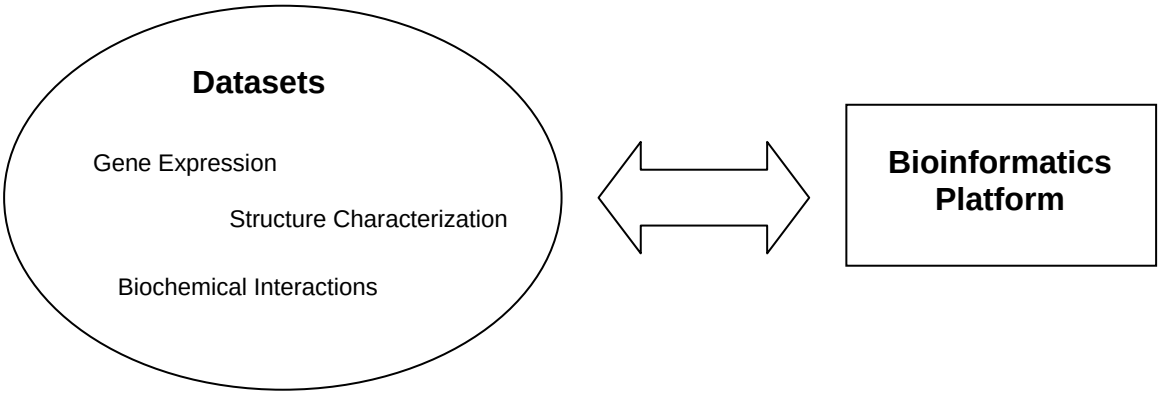
enter into the study of glycomics, I would recommend that the same audience read any follow-on papers by Raman et al. It is likely that the steps toward integration will be more advanced in the future and perhaps give rise to more complex and interesting questions to study.

Looking to the future, if I were to design a project to further the advancement of glycomics research, I would focus on improving the bioinformatics resources available to decentralized researchers. One large issue is that there is no generally agreed-upon set of standards to compare glycan structures and medium to exchange this data. Thus, my project would examine the best methods for improving bioinformatics resources for glycomics. The improvements I would explore in my project would not be the creation of a new descriptive language or the testing of the most effective algorithm in automated sequence identification. Rather, my project would be an in-depth look at all the languages and algorithms currently being employed, and would pool that information together into a paper.

The nature of my project proposal would best be characterized as a research-oriented review paper. It is similar to a review paper in the sense that my final project will be a review on the state of bioinformatics as it pertains to glycomics. However, due to the evolving nature of the platform in the field of glycomics, the review will not be rehashing past reviews. There is a tremendous amount of research and improvements being undertaken in this field and new papers are continually being published: what is deemed useful, necessary, or obsolete may change over the course of a year or a summer (Goetz, 774). One purpose of bioinformatics is to integrate information from decentralized sources, and a major obstacle facing glycomics is how to normalize the linking of all this information (Lutteke, 1). If all the researchers had a better understanding of what descriptive languages were being used, the standards used to characterize a glycan, and other universal benchmarks, the pace of advancement in this field would surely

increase (Lutteke, 3). That is why my project would be very relevant. It would help condense the current methods and standards so that more review papers could be published that are even better, more current representations of an increasingly organized structure. **Figure 1** depicts what Raman discussed in his review paper as well as what my project hopes to address:

Figure 1



The datasets in **Figure 1** represent the technological areas that generate important datasets that need to be integrated in order for a true understanding of glycans to be achieved. The bioinformatics platform as described by Raman, ties these datasets together. The double arrow represents the dynamic and synergetic nature between the two areas that will further glycan research. The better our datasets become, the more relevant and useful an effective bioinformatics platform would be. Likewise, the better the bioinformatics platform, the more quickly researchers will be able to identify relationships between datasets and will be able direct and improve that aspect of glycan investigation. My project will go deeper into the bioinformatics aspect of this dynamic relationship when compared to Raman, but will be similar in that it will be a source for other researchers to understand where the field is currently at and what should be done to improve upon it.

In order to determine the ideal environment to advance glycomics, multiple areas would need to be addressed. Ideally, the project will find trends as to what is useful for current researchers. Whether it is the increased employment of the XML language as a result of data heterogeneity or the increased reliance on various aspects of the chemical makeup of a glycan for characterization, the project aims to identify the glycomics landscape (Cheung, 15). There is also the possibility that the currently employed techniques and standards are not converging but rather remaining separated. If this is the case, my project will identify why, and propose what steps can be taken in order to help normalized and link everything. As mentioned before, my project will no doubt have characteristics like that of a review paper, but it will have a heavy focus on research and identifying trends in this field of research. Although this task appears difficult due to its broad nature, by following a general guiding framework, I believe I will be successful in my aims.

I would first outline how the study of genomics and proteomics has utilized bioinformatics tools in order to advance knowledge in each of those fields. For example, it is currently accepted that the algorithms used in genomics are not very applicable to automated MS spectrum identification in glycomics, thus necessitating further research in this area (von der Lieth, 175). Since the goal of my project would be to identify how to improve the glycomics bioinformatics platform, information as to where bioinformatics differs in relation to genomics' and proteomics' usages of bioinformatics tools must be clearly shown (von der Lieth, 165).

Once this broad view of the landscape is clarified, I would focus on proposing possible characteristics of glycans that could be compared across databases. The gene expression, structure, and biochemistry aspects of a glycan are all important datasets to compare, and my project would be to look at which datasets and what information in those datasets are being

studied most prominently. For example, with regards to structure, there are various challenges such as how glycans exhibit heterogeneous structures bound to a single protein as well as the formation of non-linear, branched structures. My proposal will identify the current state of research on these challenges and observe how solutions vary or relate (von der Lieth, 175). Other areas to look at would be the aspects of glycans that databases are currently storing. Descriptions such as glycan type, linkage, and configuration in one database could be compared to how glycans are described in another database (Cooper, 332).

Finally, I would try and identify how the Extensible Markup Language (XML) is being employed to further bioinformatics resources for glycomics. I would identify the main types of descriptive languages being used for the exchange and storage of glycan information and note how each differ and what advantages each method confers for the user. Kickuchi et al. describes in his paper the carbohydrate sequence markup language (CabosML) and how this XML description enables the description of multiple glycan characteristics; the description of repeating and cyclic structures is one example (Kikuchi, 1717). By surveying the landscape of descriptive languages, I will be able to clarify what the common descriptions are and provide an outline as to what researchers are focusing on in hope of enabling more standardized descriptions of glycans.

Glycomics research will undoubtedly expand and advance in the coming years. The bioinformatics platform supporting all of the new datasets will eventually evolve to accommodate researchers' needs; however, the speed at which it will be developed is uncertain. By continually reviewing, researching, and communicating the standards being employed, the development of bioinformatics resources for glycomics will hopefully be expedited.

Resources

Cheung, Kai. “Intersection of Semantic Web and Life Sciences.” *MBB 452 Handout* (2005): 1-65).

Cooper, Catherine et al. “GlycoSuiteDB: a new curated relational database of glycoprotein glycan structures and their biological sources.” *Nucleic Acids Res.* 29 (2001): 332-335.

Goetz, T., and C.W. von der Lieth. “Pubfinder: a tool for improving retrieval rate of relevant PubMed abstracts.” *Nucleic Acids Res.* 33 (2005): W774-774.

Kikuchi, N. et al. “The carbohydrate sequence markup language (CabosML): an XML description of carbohydrate structures.” *Bioinformatics* 21 (2005): 1717-1718.

Lutteke, T. et al. “GLYCOCIENCES.de: An Internet Portal to Support Glycomics and Glycobiology Research.” *Glycobiology* (Oct. 2005): 1-14 (ePublication).

Raman, Rahul et al. “Glycomics: an integrated systems approach to structure-function relationships of glycans.” *Nature* 2.11 (2005): 817-824.

Von der Lieth, C.W. et al. “Bioinformatics for glycomics: status, methods, requirements and perspectives.” *Brief. Bioinform* 5 (2004): 164-178.