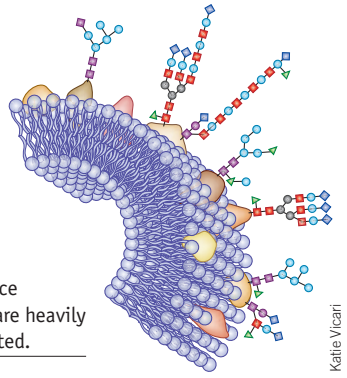
Glycoproteomics

Methods for tackling the enormously complex glycoproteome are sorely needed.

Technologies for profiling the proteome and even simple post-translational modifications such as phosphorylation are approaching maturity, but the most abundant post-translational modification, glycosylation, still remains practically unexplored at the proteome scale. This is not for lack of interest but because of a dearth of methods for profiling the enormously complex glycoproteome.

Glycans, complex chains of sugars, are not just energy-storage molecules; the important specific biological roles of protein glycosylation are being increasingly brought to light. Eukaryotic cell-surface proteins are often heavily glycosylated, indicating the importance of these modifications in cell signaling, cell-cell interactions and the immune response, for example.

There are several methodological issues that make tackling the glycoproteome particularly challenging. As a post-translational process, glycosylation is by definition nontemplated. However, unlike simple post-translational modifications such as phosphorylation, the great diversity of glycan structures makes their analysis exponentially more difficult. One single protein can have tens to hundreds of different glycan attachments. Glycosylated forms of proteins are often found in low abundance in the cell, and the modifications themselves in low stoichiometry.



Cell-surface proteins are heavily glycosylated.

Katie Vicari

To date, the glycomics field and the proteomics field have often existed in separate spheres. Glycomics researchers profile glycan structures but ignore the proteins from which they came, and proteomics researchers profile proteins while ignoring the appended glycans. However, the importance of integrating the analyses is being realized; proteomics researchers have recently reported high-throughput methods to detect protein glycosylation sites—though not the glycan structures. Glycomics researchers have been able to characterize all glycans found on single proteins—but not in high throughput.

Mass spectrometry, already a proven technology for proteomics, is likely to also be key for glycoproteomics. The very different chemistries of protein chains and glycan chains present a sequencing challenge, but high-resolution mass spectrometry instruments and newer fragmentation methods are likely to facilitate such analyses. Methods for isolating and separating glycoproteins before mass spectrometry analysis, and bioinformatics approaches for analyzing the complex data resulting from such experiments, are also needed.

We hope to see an abundance of new methods for high-throughput glycoproteomics in the near future. Making sense of the data generated from such approaches to sort out the functional roles of glycosylation, however, will take much longer.

Allison Doerr

Causal mutations in a haploid landscape

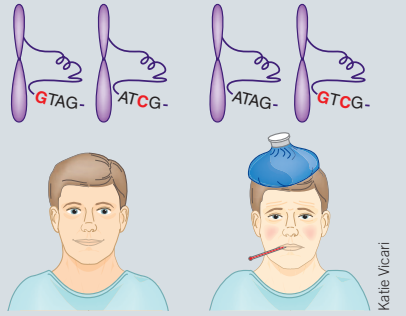
Sequencing a haploid genome and understanding the impact of its variants requires technical and computational improvements.

The rapid evolution of high-throughput sequencing has shifted the focus from data acquisition to data interpretation. Researchers working on human genomes have made big strides in primary-sequence analysis; they can map short reads, call single-nucleotide polymorphisms, insertion-deletions, or indels, and even map more complex structural variants. But two key challenges remain. The first challenge is to determine the haplotypes—the individual sequences of each chromosome in an individual's genome—especially when no information on the genomes of close relatives exists; the second challenge is to determine the functional effect of all the detected variants and to pinpoint those that cause disease.

The year 2011 has seen several efforts to determine an individual's haplotype-resolved genome, for example, the high-throughput sequencing of fosmid clones in the Max Planck One genome (*Genome Res.* **10**, 1672–1685; 2011) or the sequence analysis of individual chromosomes separated via a microfluidic device (*Nat. Biotechnol.* **29**, 51–57; 2011). These studies showed the occurrence of novel variants in many genes, and they underscored the importance of phasing these mutations to be able to assess the impact they can have on the individual.

Assigning mutations to a haplotype is only the first step; one then needs to decide which of these mutations are functionally deleterious to the individual—a task Gregory Cooper and Jay Shendure referred to as "finding needles in a pile of needles once the haystack has been cleared away" (*Nat. Rev. Genet.* **12**, 628–640; 2011). Computational approaches based on evolutionary conservation of sequences, biochemical properties of protein sequences and structural information have come a long way in providing candidate lists of potential mutations both in protein-coding sequences as well as in noncoding stretches of the genome. But even at their best, these programs only provide candidate lists. They will need to be complemented with large-scale experimental approaches to analyze these variants and provide molecular phenotypes that lead to the functional assessment of a given mutation.

Nicole Rusk



Katie Vicari

The impact of mutations may depend on whether they occur together on the same chromosome or not.