4.1

Cancer proteomics — connecting genotype with molecular phenotype

Janne Lehtiö[1]

[1]Karolinska Institutet, Science for Life Laboratory, Stockholm, Sweden

The explosion of genomics data has improved our understanding of cancer greatly in recent years. However, the knowledge of how genomic aberrations affect the functional proteome at the systems level is still very limited. Proteome data represents the combined effect of epigenetic, transcriptional and translational regulation and will therefore provide an important molecular phenotype data layer for multi-omics analysis. To allow effective systems biology analysis including proteomics, we have generated tools that take advantage of massive genomics data by incorporating sequence information to the proteomics data-analysis pipeline. This will allow protein level analysis of gene variants as well as detection of novel protein coding regions. To control error rate in variant detection, we have a combined experimental isoelectric point data from peptide fractions (HiRIEF LC-MS/MS) and bioinformatics approaches into the proteogenomics workflow (IPAW). A proteogenomics analysis of histologically human tissues using the IPAW pipeline reveals novel coding regions. When applied on breast cancer tumor sample, we could demonstrate in-depth quantitative analysis revealing drug target interesting correlations as well as discovers putative cancer neoantigens. Further, to narrow down on proteins involved in immediate cellular response to drug treatment, we have analyzed cell subcellular relocation of protein post drug treatment using SubCellBarcode based proteome wide location analysis.

4.2

The relation between infection, autoimmune mechanisms and Parkinson's disease

Michel Desjardins[1]

[1]Département de pathologie et biologie cellulaire, Université de Montréal

Parkinson's disease (PD) is a neurodegenerative disorder caused by the progressive loss of dopaminergic neurons (DNs). While it affects close to 3% of the population over 75 years of age, a significant proportion of PD patients, possibly as high as 10%, develops the disease due to familial, transmitted mutations, at a much earlier age.Two of the genes mutated in early-onset PD, PINK1 and Parkin, are involved in mitophagy, the process by which damaged mitochondria are captured for recycling within autophagosomes. Hence, it is assumed that in the absence of PINK1 or Parkin, failure to eliminate non-functional mitochondria in DNs results in the accumulation of toxic organelles, excessive oxidative stress and cell death. Although compelling, this model has proven difficult to validate in vivoas there is, so far, little evidence for a deregulation of mitophagy within DNs in PD. Furthermore, Parkin- and PINK1-independent pathways of mitophagy exist, suggesting the involvement of these proteins in PD through different mechanisms. Importantly, PINK1 and Parkin KO mice are generally healthy and display no signs of the disease. We have shown recently that PINK1 and Parkin play a role in the immune system by inhibiting the formation of mitochondria-derived vesicles (MDVs) and mitochondrial antigenpresentation, a process we refer to as MitAP. In the absence of PINK1/Parkin, stress conditions such as inflammation induced by LPS treatment activate MitAP in antigen presenting cells (macrophages and dendritic cells) in vivo, a process leading to the establishment of autoreactive CD8+ T cells. LPS being the major component of the outer membrane of Gram-negative bacteria, we went on to show that gut infection with enteropathogenic E. coli (EPEC) induces MitAP and the elicitation of anti-mirochondrial CD8+ T cells in PINK1 KO mice. Remarkably, these animals display severe motor impairment as early as 3 months after infection, reversible by L-DOPA treatment. The link between infection, autoimmune mechanisms and the emergence of parkinsonism in PD-susceptible mice opens novel avenues for the development of therapeutics.

4.3

Identification of mechanisms of activity and resistance to thalidomide analogs with a targeted quantitative immuno-mass spectrometry assay

Adam S. Sperling[1,2], Michael Burgess[2], Hasmik Keshishian[2], Jessica A. Gasser[1,2], Max Jan[1,2], Mikolaj Slabicki[1,2], Peter G. Miller[1], Rohan Sharma[1], Dylan N. Adams[1], Mariateresa Fulciniti[1], Namrata D. Udeshi[2], Eric Kuhn[2], Nikhil C. Munshi[1], Steven A. Carr[2], Benjamin L. Ebert[1,2]

[1]Dana-Farber Cancer Institute, [2]Broad Institute of MIT and Harvard

Pharmacologic agents that modulate ubiquitin ligase activity to induce protein degradation are a major new class of therapeutic agents. We developed a high-throughput, quantitative, targeted mass spectrometry (MS) assay leveraging immune enrichment of specific peptides to measure the levels of proteins that are degraded by the CRL4^{CRBN} ubiquitin ligase in the presence of thalidomide analogs. Using this immuno-MS assay to determine the levels of eight protein substrates, we defined key differences in substrate specificity between thalidomide derivatives, characterized distinct kinetics of degradation for different substrates, and identified a novel mechanism of resistance to this class of drugs mediated by competition between substrates for access to the ubiquitin ligase. We demonstrated that increased expression of a non-essential substrate can lead to decreased degradation of other substrates that are critical for anti-neoplastic activity of the drug, resulting in drug resistance. These findings suggest that tissue-specific activity of drugs that induce protein degradation will depend on the levels of the ubiquitin ligase as well as the expression of substrates. The quantitative mass spectrometry assay we describe is a powerful tool to characterize the activity of novel molecules that induce protein degradation, to evaluate the activity of such molecules in vitro and in vivo, and to predict sensitivity and resistance to this class of therapeutic agents.

4.4

Crosslinking mass spectrometry and single particle cryoEM describe the structure of a novel translocon in complex with the ribosome

Michael J. Trnka[1], Philip T. McGilvray[2], Robert J. Keenan[2], Alma L. Burlingame[1]

[1]University of California San Francisco, CA 94158, [2]University of Chicago, IL 60637

Multi-pass transmembrane proteins play key roles in numerous aspects of cell physiology. These proteins are synthesized at the endoplastic reticulum (ER) and their insertion, folding, and assembly into the membrane is coordinate by the 'translocon', a poorly defined and dynamic assembly comprising the Sec61 translocation channel and a variety of accessory subunits that act cotranslationally with the ribosome. We have isolated a novel eukaryotic, 93-protein ribosome-translocon complex (RTC) that facilitates the biogenesis of ~20% of all multi-pass membrane proteins. The structure of this RTC was characterized by single particle EM reconstruction supported by crosslinking mass spectrometry (CLMS).

Over 1200 unique crosslinked residue pairs were identified at an FDR of 0.5%. A subset of 130 crosslinks localized the translocon at the exit channel of the ribosome and facilitated modeling of electron density corresponding to the lumenal and membrane subunits of the translocon. Crosslinking was performed using the membrane soluble reagent DSS, and analysis was performed using sequential high resolution HCD and ETD product ion scans of the same precursor ion, on an Fusion Lumos mass spectrometer. The decoy distribution was modeled using 930 randomized sequences, corresponding to 10 decoy versions of each target protein. FDR calculations were adjusted for the discrepancy in database sizes using a mathematical approach described here. Additional checks on the validity of the dataset came from agreement of the ribosomal crosslinks with high-resolution structure and by monitoring ribosomal aggregation during reaction optimization with negative stain EM. We discuss the relationship between crosslink violation rate and FDR.

4.5

Systematic profiling of HLA class I immunopeptidome improves neoantigen binding prediction

Susan Klaeger[1]

[1]Broad Institute

Highly polymorphic class I HLA proteins present short peptides from endogenous or foreign proteins to cytotoxic T cells. Each allele is estimated to present 1,000–10,000 peptides, however the rules of antigen presentation are not fully understood. Mass spectrometry allows for direct identification of endogenously processed and presented peptides. Using a single allele expressing cell line, the underlying criteria for antigen presentation can be systematically studied. This information is of high value for training epitope prediction models used in e.g. personalized vaccine generation. Our strategy improves the performance of current predictive algorithms and provides a rapid and scalable method to generate rules for the substantially diverse set of human HLA alleles.

We developed a mono-allelic MS approach to profile endogenously presented HLA-peptides, whereby the HLA class I deficient B721.221 cell line expresses a single allele of interest and eluted HLA peptides are analyzed by LC-MS/MS.

Using this approach, we have generated binding data for 95 HLA- A, B, C and G alleles identifying more than 200,000 peptides and covering the most frequent alleles in the population. This extensive dataset enables peptide-binding and proteasomal cleavage motifs to be elucidated on a single allele basis. HLA- A and B alleles present more peptides of length 10–11 than C alleles, while C alleles have a higher propensity for 8-mers. Correlation-based analysis of binding motifs revealed that HLA-A and B motifs are more specific whereas C motifs are less stringent and thus share more overlapping binders. This data is used to train neural network models to predict potential MHC-binders from genomic data. We show superior performance in discriminating tumor-presented epitopes than state of the art predictors. The predictive value of the models is determined using MHC I peptide binding data generated by LC-MS/MS from primary tumor samples and by performing targeted analyses for specific peptides of interest.

This work vastly expands the collection of endogenous HLA-allele specific peptides, which not only allows biological insights into the principles of epitope presentation but also improves prediction algorithms for endogenous neoantigen selection for tumor vaccines.

4.6

Ribosome profiling predicts novel unannotated open reading frames that contribute peptides to the MHC class I immunopeptidome

Tamara Ouspenskaia[1], Travis E. Law[1], Karl R. Clauser[1], Susan Klaeger[1], Derin B. Keskin[4], Bo Li[1], Elena Christian[1], Phuong M. Le[4], Zhe Ji[5], Wandi Zhang[4], Pavan Bachireddy[4], Siranush Sarkizova[2], Nir Hacohen[1,3], Steven A. Carr[1], Catherine J. Wu[1,4], Aviv Regev[1]

[1]Broad Institute of MIT and Harvard, Cambridge, MA, [2]Harvard University, Cambridge, MA, [3]Massachusetts General Hospital, Boston, MA, [4]Dana Farber Cancer Institute, Boston, MA, [5]Northwestern University, Chicago, IL

Cancer-specific neoantigens, derived from somatic mutations, presented on the MHC class I, and recognized by the immune system, have emerged as an important target to drive immunotherapy. Currently, neoantigen predictions are based on mutations detected by whole exome sequencing, which covers a pre-determined set of *annotated* protein-coding genomic regions, and often falls short for patients with few somatic mutations.

Ribosome profiling (Ribo-seq) has suggested a plethora of translated novel unannotated open reading frames (nuORFs). We hypothesized that nuORFs can provide another source of neoantigens in cancer. We focused on nuORFs in two categories: 1) expressed in healthy and cancer cells, and have acquired cancer-specific somatic mutations; 2) upregulated in or specific to cancer cells.

We performed Ribo-seq on primary healthy and cancer cells and cell lines from melanoma, glioblastoma, colon carcinoma and chronic lymphocytic leukemia and built a database of predicted nuORFs. We performed MHC class I immunoprecipitation and LC-MS/MS on the same samples. We also searched our collection of mono-allelic MHC class I immunopeptidome MS spectra from 92 common HLA alleles against our pan-tissue nuORF database. We found HLA-presented peptides derived from thousands of nuORFs, found within 5′ and 3′ UTRs, lncRNAs, pseudogenes, and other RNA species.

To identify tumor-specific mutations in nuORFs, we performed whole genome sequencing on patient-matched healthy and cancer cells and mapped somatic mutations to nuORFs. To identify nuORFs upregulated in or specific to cancer cells, we compared translation levels of nuORFs between healthy and cancer cells of the same origin. We found translated nuORFs with cancer-specific mutations and nuORFs highly upregulated in and specific to cancer cells, suggesting that they can give rise to neoantigens.

In conclusion, nuORFs are translated, detected on MHC class I, acquire somatic mutations, are expressed in tissue- and cancer-dependent manner and should be considered in the search for neoantigens.

4.7

Genomic dark matter is a major source of targetable tumor-specific antigens

Sibylle Pfammatter[1], Eric Bonneil[1], Joel Lanoix[1], Celine Laumont[1], Krystel Vincent[1], Chantal Durette[1], Jean-Philippe Laverdure[1], Mathieu Courcelles[1], Marie-Pierre Hardy[1], Sebastien Lemieux[1], Claude Perreault[1], Pierre Thibault[1]

[1]Universite de Montreal, Institute for Research in Immunology and Cancer

Neoantigens encoded tumor-specific mutated genes and represent ideal targets for cancer immunotherapy. While significant efforts have been devoted to the identification of mutated tumor specific antigens (TSAs) for cancer immunotherapy, accumulating evidences suggest that a large proportion of cancer mutations originate from non-coding regions. Here, we combined high-field asymmetric waveform ion mobility spectrometry (FAIMS) and isobaric peptide labeling to enhance the identification of peptides presented by MHC class I molecules (MHC I), and developed a novel proteogenomic approach to uncover the repertoire of TSAs coded potentially from all genomic regions. We applied these methods to the discovery of TSAs from B-lymphoblastoid cells and from human leukemia primary tumors. In cancer cell lines and human primary tumors, we identified more than 40 TSAs, of which 90% derived from allegedly noncoding regions that are typically inaccessible by standard exome-based approaches. The possibility that several of these antigens could be shared by multiple tumor types open up interesting in cancer immunotherapy.

4.8

Proteogenomics and immunopeptidomics for the development of personalized cancer immunotherapy

Michal Bassani-Sternberg[1,2], Chloe Chong[1,2], Markus Müller[3], HuiSong Pak[1,2], Florian Huber[1,2], Justine Michaux[1,2], Brian Stevenson[3], Julien Racle[4], David Gfeller[2,3], George Coukos[1,2]

[1]University Hospital of Lausanne, Lausanne, Switzerland, [2]Ludwig Institute for Cancer Research Lausanne, [3]Swiss Institute of Bioinformatics, Lausanne, Switzerland, [4]University of Lausanne, Lausanne, Switzerland

The remarkable clinical efficacy of the immune checkpoint blockade therapies has motivated researchers to discover immunogenic epitopes and exploit them for personalized vaccines and T cell based therapies. Mutated human leukocyte antigen binding peptides (HLAp) are currently the leading targets. We and others have shown that the direct identification of tissue-derived immunogenic neoantigens by mass spectrometry (MS) is feasible. However, most studies attempt to identify neoantigens based on HLA binding prediction tools. We have compiled a large immunopeptidomics database across dozens of HLA allotypes. By taking advantage of co-occurring HLA-I alleles, we rapidly and accurately identified HLA-I binding motifs. We have shown that training HLA-I ligand predictors on refined motifs significantly improves the identification of neoantigens. Recently, we have acquired the largest reported HLA-II immunopeptidomics dataset. We introduced novel algorithmic tools to analyze this data and developed for the first time HLA-II epitope prediction tool trained on immunopeptidomics data that results in major improvements in prediction accuracy.

In contrast to the private neoantigens, tumor-specific antigens that are shared across patients may be more attractive for immunotherapy. Recent studies have focused on the discovery of aberrantly-expressed non-canonical antigens, which expands the repertoire of targetable epitopes through the translation and presentation of presumably non-coding regions. However, their identification requires highly sensitive and accurate MS-based proteogenomics approaches.

We have developed a novel analytical pipeline that can precisely characterize the non-canonical HLAp repertoire. The workflow incorporates whole exome sequencing, both bulk and single cell transcriptomics, ribosome profiling, and a combination of two MS/MS search tools with group-specific false discovery rate calculations for accurate HLAp identification. We identified more than 400 shared and tumor-specific non-canonical HLAp derived from the expressed lncRNAs, transposable elements and alternative open reading frames. Moreover, non-canonical HLAIp were experimentally confirmed to be shared across tumors through targeted MS, by which synthetic heavy isotope-labelled peptides were spiked into the peptidomic sample. This analytical platform holds great promise for the discovery of novel cancer antigens for cancer immunotherapy.