Computational and experimental approaches to determine the immuno-peptidome landscape on cancer cells

Georges Bedran[1,2], Marcos Yebenes[1,2], Sachin Kote[1,2], Irena Dapic[1,2], Kamila Pawlicka[2], Satya Saxena[1,4], David Goodlett[1,3], Robin Fahraeus[1,2], Ted Hupp[1,2], Javier Alfaro[1,2]

[1]International Centre for Cancer Vaccine Science, University of Gdansk, Gdansk, Poland

[2]CRUK, University of Edinburgh, Edinburgh, United Kingdom

[3]University of Maryland, Baltimore, United States

[4]Deurion LLC, Seattle, WA, United States

1. Introduction

Throughout the years vaccines have been demonstrated to be one of the cheapest and most effective methods to prevent and/or treat infectious diseases. The use of proteogenomic approaches [1] applied to the identification of antigens has the potential of establishing protein arrays that may accelerate the identification of vaccine candidates. The focus of this work is on establishing the best computational workflows for neo-epitope identifications using a combination of genomics, transcriptomics and proteomics datasets.

Figure 1: A schematic presentation of a proteogenomic workflow.

2. Data exploration

Number of epitopes identified per patient

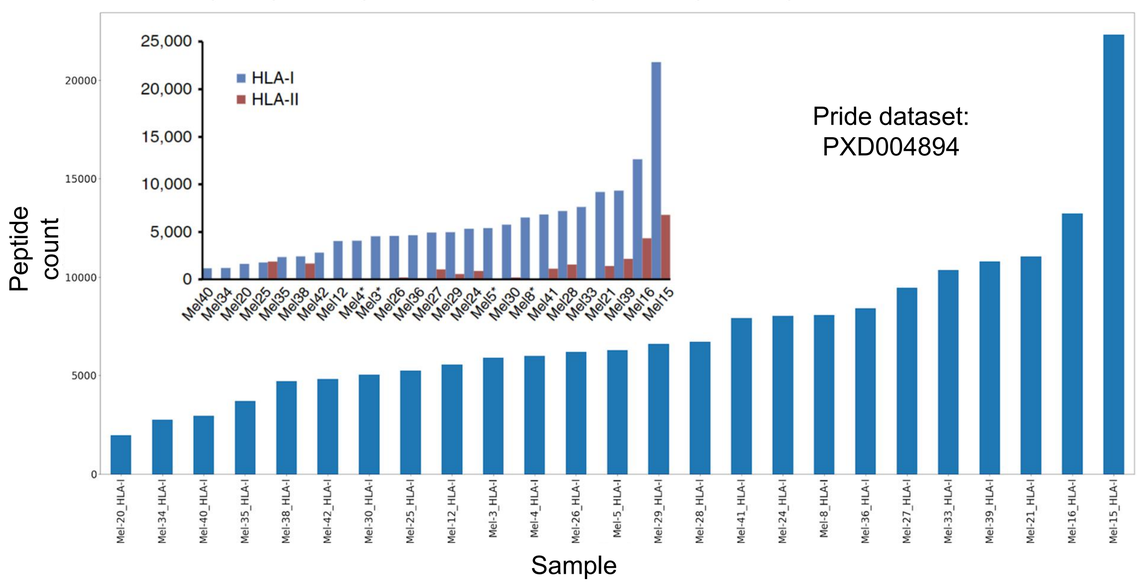

Figure 2: Barplot showing the number of HLA-I peptides detected per patient using our strategy versus the published results.

Post Translational Modifications

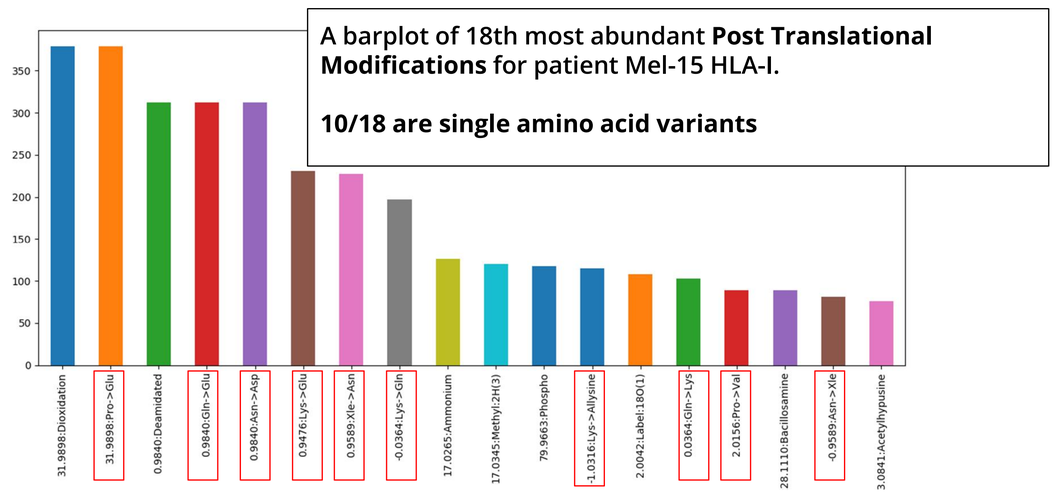

Figure 3: Barplot showing the most frequent PTMs demonstrating the capability of identifying single amino acid variations.

3. Unraveling the source

Are there hotspots of antigen presentation ?

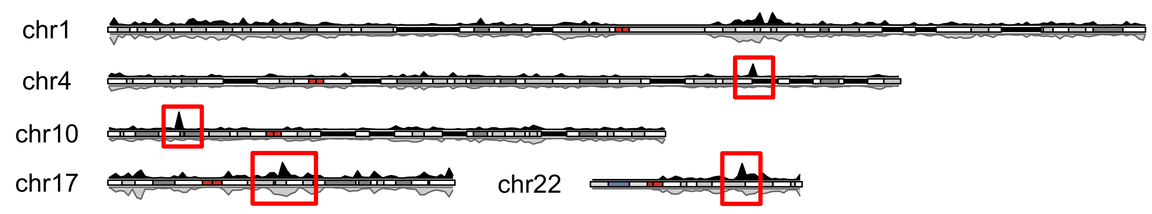

Figure 4: An illustration showing the densities of the identified HLA-I peptides and genes over the karyotype in black (top panel) and grey (bottom panel) respectively.

Overall the regions with high gene densities are in concordance with a high antigenic expression. However, some regions shown in the figure above do not follow this trend. That's why we suspected having hotspots of antigen presentation.

Projection to 2D space

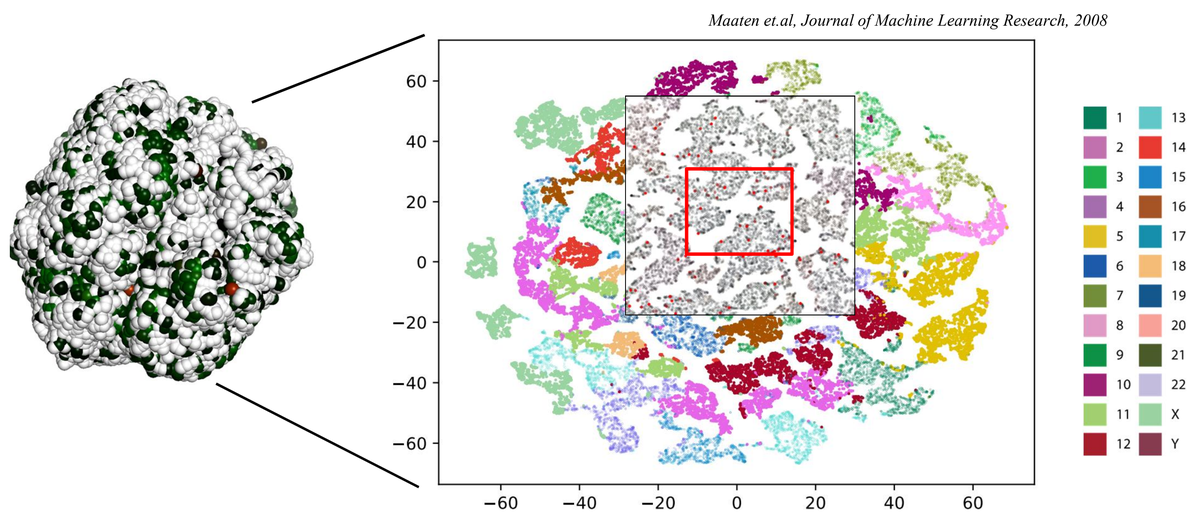

Figure 5: A Tsne [2] projection of the genome's 3D structure in 2D space that preserves the features' geo-proximity. Therefore, allowing to visualize hotspots of antigen presentation.

Tan et. al [3] were able to reconstruct the genome structures of single diploid human cells from a lymphoblastoid cell line. We used one of their 3D structures, as a prototype, in order to highlight regions of antigen presentation to account for potential spatially close chromosomes' hotspots.

5. Acknowledgements

We would like to thank the FNP (Foundation for Polish Science) for funding this project. In addition, all the calculations were possible thank to the PL-Grid infrastructure.

References

[1] J. A. Alfaro, A. Sinha, T. Kislinger, and P. C. Boutros, "Onco-proteogenomics: cancer proteomics joins forces with genomics," *Nature Methods*, vol. 11, p. 1107, Oct. 2014.

[2] L. Maaten and G. Hinton, *Visualizing High-Dimensional Data using t-SNE*, vol. 9. Jan. 2008.

[3] L. Tan, D. Xing, C.-H. Chang, H. Li, and X. S. Xie, "Three-dimensional genome structures of single diploid human cells," *Science*, vol. 361, p. 924, Aug. 2018.

Contact

Author : Georges Bedran

Email : georges.bedran@phdstud.ug.edu.pl

European Funds Smart Growth | FNP Foundation for Polish Science | European Union European Regional Development Fund