

# Resumen

El bloqueo de los puntos de control como CTLA-4 y PD-1/PD-L1 ha demostrado la actividad de efectores inmunes antitumorales, por lo que ha aumentado el interés en el desarrollo de inmunoterapias contra el cáncer. Las células T tienen un rol fundamental ya que son capaces de reconocer y eliminar células tumorales que expresan antígenos. Por esta razón, identificar a los antígenos de cáncer resulta central. En su evolución, los tumores acumulan mutaciones somáticas, las cuales son capaces de generar un tipo de antígenos denominados neoantígenos, los cuales son altamente específicos. El desarrollo de técnicas de secuenciación de nueva generación y bioinformáticas ha permitido obtener información genómica de tumores e identificar estos potenciales neoantígenos. Sin embargo no todas las mutaciones detectadas resultan inmunogénicas y es importante acotar la búsqueda. La presente tesis explora técnicas computacionales que permiten priorizar neoepítopes inmunogénicos.

Se seleccionaron pacientes con melanoma y otros tumores sólidos que recibieron distintos tipos de inmunoterapia. Se les realizó secuenciación exómica pareada de muestras sanas y tumorales, así como del transcriptoma tumoral, en caso de tener muestra suficiente. En base a estos datos ómicos se identificaron mutaciones somáticas localizadas en regiones codificantes y se predijeron péptidos capaces de presentarse en el MHC de los pacientes. La cantidad de mutaciones somáticas fue elevada en algunos pacientes, resultando en un gran número de neoepítopes candidatos. Por esta razón, se exploraron variables adicionales que permitieran priorizar adecuadamente a los neoepítopes verdaderos. Entre ellas evaluamos características relacionadas con el procesamiento y presentación antigénicos, la expresión de transcriptos mutados, la capacidad de tolerización inmune por similitud a la proteína *wild type* y la composición aminoacídica de los neopéptidos, entre otros. Adicionalmente, se desarrollaron modelos predictivos basados en aprendizaje automático con el propósito de priorizar a los neoepítopes. La inmunogenicidad de los candidatos fue evaluada utilizando muestras biológicas de pacientes en tratamiento con inmunoterapia, con ensayos de ELISpot y de multímeros de MHC con códigos de barras. Finalmente, trasladamos el conocimiento capitalizado para identificar epítopes contenidos en vacunas tumorales celulares.

Como resultado de este trabajo de tesis, primero logramos establecer un *pipeline* bioinformático capaz de identificar neoepítopes en pacientes de cáncer. Extendiendo este *pipeline* a nuevos pacientes fue posible caracterizar nuevos neoepítopes inmunogénicos, así como neopéptidos no inmunogénicos. En base a este conjunto de datos, demostramos que las herramientas computacionales disponibles que habían sido desarrolladas hasta el momento tenían un desempeño muy limitado en la priorización de neoantígenos. Posteriormente desarrollamos un modelo capaz de mejorar la detección de neoepítopes, el cual reveló que la composición aminoacídica es la característica de mayor influencia en la inmunogenicidad, y que

características inferidas del microambiente tumoral están asociadas con el reconocimiento inmune de neopeptidos. Por otro lado, adaptamos el *pipeline* bioinformático para identificar antígenos presentes en líneas celulares tumorales que componen VACCIMEL, una vacuna terapéutica para melanoma cutáneo. Este reveló que las vacunas celulares son capaces de estimular respuestas contra una gran cantidad de aloantígenos y antígenos de cáncer, entre los cuales encontramos un número importante de neoantígenos.

En general, afirmamos que la presentación antigénica estimada con métodos como NetMHCpan es aún el mejor parámetro para realizar predicciones de neoepitopes, y que la mejora de modelos predictivos adicionales requiere datos de alta calidad. En este sentido, una contribución importante de esta tesis es la caracterización de nuevos neoepitopes inmunogénicos. En particular, en el Artículo IV hemos identificado neoepitopes derivados de mutaciones *drivers* y frecuentes en melanoma que pueden emplearse en estrategias inmunoterapéuticas semipersonalizadas.

## Palabras clave

Neoepitope, Inmunoterapia, Inmunoinformática



## Summary

Immune checkpoint blockade of CTLA-4 and PD-1/PD-L1 has proven the immune system potential for antitumor effects, increasing the interest in cancer immunotherapies. T cells play a fundamental role since they are capable of recognizing and eliminating cells expressing tumor antigens. For this reason, identifying cancer antigens is paramount. In their evolution, tumors accumulate somatic mutations capable of generating a type of antigens called neoantigens, which are highly specific. The development of next-generation sequencing and bioinformatic techniques has made it possible to extract genomic information from tumors and identify these potential neoantigens. However, not all mutations are immunogenic and it is important to narrow the search. The present thesis explores computational techniques for prioritizing immunogenic neoepitopes.

Patients with melanoma and other solid tumors who received different types of immunotherapy were selected. Paired exome sequencing of healthy and tumor samples was performed, as well as tumor transcriptome sequencing if there was enough sample. Based on these omics data, somatic mutations located in coding regions were identified, and peptides capable of being presented by the MHC of the patients were predicted. The number of somatic mutations was high in some of the patients, resulting in

a large number of neoepitope candidates. For this reason, additional variables that would appropriately prioritize true neoepitopes were explored. Among them, we evaluated features related to antigenic processing and presentation, the expression of mutated transcripts, the capacity for immune tolerization due to similarity to the wild-type peptide, and the amino acid composition of neopeptides, among others. In addition, we developed predictive models based on machine learning to prioritize the neoepitopes. The immunogenicity of the candidates was evaluated with ELISpot or barcoded MHC multimer assays, using biological samples obtained from patients receiving immunotherapy. Finally, we transferred the capitalized knowledge to identify epitopes contained in cellular tumor vaccines.

As a result of this thesis, first we established a bioinformatics pipeline capable of identifying neoepitopes in cancer patients. By extending the pipeline to new patients, it was possible to characterize new immunogenic neoepitopes, as well as non-immunogenic neopeptides. Based on this data set, we demonstrated that available computational tools that had been developed so far had very limited performance in prioritizing neoantigens. Subsequently, we developed a model capable of improving the detection of neoepitopes, which revealed that the amino acid composition is the feature with the greatest influence on immunogenicity, and that inferred tumor microenvironment's properties are associated with immune recognition of neopeptides. On the other hand, we adapted the bioinformatics pipeline to identify antigens present in tumor cell lines that make up VACCIMEL, a therapeutic vaccine for cutaneous melanoma. This work disclosed that cellular vaccines are capable of stimulating responses against multiple alloantigens and cancer antigens, among which we find a significant number of neoantigens.

In general, we state that antigen presentation estimated with methods like NetMHCpan is still the best parameter to perform neoepitope prediction, and that further improved predictive models require high-quality data. In this sense, one important contribution of this thesis is the characterization of novel immunogenic neoepitopes. Particularly, in Article IV we have identified neoepitopes derived from driver and frequent mutations in melanoma that can be applied in semi-personalized immunotherapeutic strategies.

## Keywords

Neoepitope, Immunotherapy, Immunoinformatics

