

Neoantigen Design and Development for Personalized Cancer Vaccines

Subjects: Oncology

Contributor: Tiantian Zhang, Esaw Kurban, Zhe Wang

Neoantigens, also known as tumor-specific antigens, are novel antigens originating from tumor-specific alterations such as genomic mutations, dysregulated RNA splicing, and post-translational modifications. Neoantigens, recognized as non-self entities, trigger immune responses that evade central and peripheral tolerance mechanisms. With the notable strides in cancer genomics facilitated by next-generation sequencing technologies, neoantigens have emerged as a promising avenue for tumor-specific immunotherapy grounded in genomic profiling-based precision medicine.

Keywords: neoantigens ; immunotherapy ; sequencing ; mutations ; neoantigen prediction ; prioritization ; validation ; cancer

1. Introduction

The field of cancer research has witnessed remarkable advancements over the past few decades, unveiling intricate dynamics between malignant cells and the immune system. Amid the myriad of novel avenues that have surfaced, the exploration of neoantigens has emerged as a captivating frontier within cancer immunotherapy. Neoantigens, originating from somatic mutations within the tumor genome, possess an unparalleled capacity to trigger precise immune responses, potentially reshaping the landscape of personalized cancer treatment.

At the core of virtually all immunotherapeutic strategies lies the induction and activation of tumor-specific T cells. Neoantigens, aptly named, represent the distinctive epitopes that emerge from modified gene products and novel proteins resulting from mutations within the genome's coding regions. These neoantigens are presented for recognition by T cells after being processed ^[1]. Compared to various tumor-associated antigens (TAAs), neoantigens provide a distinct advantage. While TAAs exhibit elevated levels on tumor cells but are also expressed at lower levels on healthy cells, neoantigens are expressed in the tumor tissues and absent in the normal tissues. Given that TAAs remain non-mutated self-antigens, their recognition can be hampered by central T cell tolerance mechanisms, potentially accounting for the subdued T cell responses. In contrast, T cells primed to target neoantigens can circumvent the suppressive impacts of negative selection within the thymus, due to the pronounced antigenicity conferred by somatic mutations within tumors ^[2] ^[3] ^[4]. This unique attribute of neoantigens mitigates the risk of "off-target" harm to normal tissues and circumvents the constraints of central or peripheral tolerance, offering an individualized vaccine capable of stimulating the activation of tumor-specific T cells ^[5].

In the context of neoantigen-based immunotherapy, synthetically engineered neopeptides are administered to patients with the aim of triggering an immune response, particularly engaging CD8⁺ and CD4⁺ T cells to recognize the neoantigens and target and eliminate tumor cells ^[6]. The effectiveness of neoantigens hinges upon various factors, with tumor mutation burden (TMB) and the presentation and recognition of neoantigens being of paramount importance. A higher TMB is anticipated to yield a greater pool of tumor-specific antigens, enhancing the likelihood of inducing tumor antigen-specific T cells. To validate this hypothesis, a multitude of studies have been designed, stratifying immunotherapy-treated patients based on TMB levels. The outcomes consistently reveal that patients with higher TMB experience improved results, including enhanced progression-free survival and overall survival ^[7] ^[8]. Within the intricate landscape of tumor neoantigens, genetic anomalies in tumor cells—ranging from somatic point mutations and insertions to deletions and chromosomal translocations—undergo transcription and translation, ultimately giving rise to mutated peptides. These peptides subsequently undergo hydrolysis and are presented by major histocompatibility complex (MHC) molecules, facilitating their recognition by T cells ^[9]. Factors such as peptide splicing, the antigen processing and presentation machinery, and peptide affinity can influence MHC identification. Moreover, T cell recognition can be influenced by the extent of tumor infiltrating lymphocytes, culminating in the determination of the neoantigen's potential to elicit a robust immune response ^[10].

To date, a wide array of neoantigen-based vaccines have undergone evaluation in patients with various types of tumors. These vaccines encompass peptide, nucleic acid, and dendritic cell (DC) vaccine modalities. Peptide and nucleic acid vaccines primarily derive from predicted neopeptides resulting from somatic mutations, such as single nucleotide variations, frameshift insertions or deletions, and gene fusions. In contrast, DC vaccines are produced by loading dendritic cells with neoantigens. Several techniques have been employed for this purpose, including pulsing with synthetic peptides, transfection using mRNA, and pulsing with autologous whole tumor lysate ^[11].

Predicting and identifying neoantigens involves a multidisciplinary approach encompassing genomics, bioinformatics, and immunology.

2. Tumor Biopsy and Next-Generation Sequencing

Upon diagnosis of the patient, the process initiates with the selection of tumor tissue samples. These samples are meticulously curated to represent areas abundantly populated by tumor cells. Histological evaluation is of paramount importance in distinguishing cancerous tissue from healthy tissue and offers valuable insights into the tumor's histopathological attributes, encompassing its type, grade, and stage. Once histological analysis definitively confirms the presence of tumor cells, the selection of tissue becomes pivotal. Careful consideration is given to identifying tumor-rich regions for subsequent genomic and proteomic analyses. These regions are expected to harbor the specific genetic mutations and neoantigens pertinent to the patient's cancer, rendering them the focal point for personalized immunotherapies.

The tumor is biopsied, and both the cancerous and normal tissues are sequenced. Whole exome sequencing (WES), which is a powerful genomic technique, focuses on sequencing the coding regions, or exons, of genes in an individual's genome. While the majority of the human genome consists of non-coding regions, exons are where most disease-associated mutations are found. By sequencing only the exonic regions, WES enables researchers to identify single nucleotide variations, small insertions, and deletions ^{[12][13]}. Unlike WES, which focuses on the genome's DNA, transcriptome analyses, or RNA sequencing (RNA-seq) examines the transcripts that are produced from genes and serve as templates for protein synthesis. Specifically, RNA-Seq can detect alternative splicing events, where different exons are included or excluded from the final mRNA transcript and result in the production of diverse protein isoforms. Additionally, RNA-Seq captures gene fusions arising from chromosomal rearrangements, which lead to the expression of fusion transcripts that are translated into fusion proteins with unique antigenic properties. RNA-seq involves extracting RNA molecules from cells or tissues and converting them into complementary DNA (cDNA) through reverse transcription. These cDNA fragments are then sequenced to quantify the abundance of various RNA molecules ^[14]. The incorporation of RNA-Seq into mutanome analyses enables prioritization of highly expressed mutations over nonexpressed variants, resulting in a more refined selection of potential vaccine candidates.

3. Somatic Mutation Detection

Mutation calling is a computational process that follows the sequencing of DNA (WES) or cDNA (RNA-seq) to identify genetic variations or mutations within the genome. The goal is to distinguish between the naturally occurring variations and mutations that might contribute to disease. The process involves several steps, including aligning sequenced reads to a reference genome, identifying areas with discrepancies (variations), and filtering out background noise and false positives. Currently, specialized algorithms like VarScan, SomaticSniper, Strelka, and MuTect2 GATK are employed to identify somatic mutations by comparing tumor and normal sequences. They identify somatic mutations that are unique to the tumor and not present in the individual's normal tissue. Filters are applied to remove common germline variations and retain tumor-specific alterations. By comparing the tumor genome (WES) or transcriptome (RNA-seq) with the corresponding normal tissue, researchers can pinpoint mutations that have arisen during tumorigenesis ^[15].

Following mutation calling, the subsequent annotation steps are crucial for neoantigen prediction. Annotation involves assigning functional and contextual information to the identified variants using tools like ANNOVAR, Variant Effect Predictor (VEP), or SnpEff. These tools help determine whether a variant falls within a protein-coding region, its effect on the amino acid sequence, and its potential impact on protein structure and function. Subsequently, the annotated variants are further analyzed to identify mutations that generate altered peptide sequences (neoepitopes) capable of binding to MHC molecules and eliciting an immune response. The integration of mutation calling and comprehensive variant annotation lay the foundation for accurately predicting potential neoantigens that can be harnessed for personalized cancer immunotherapy strategies ^{[16][17]}.

4. Neoantigen Prediction

Epitope prediction algorithms play a pivotal role in neoantigen prediction by evaluating the likelihood that a given peptide sequence, derived from a mutated protein, will bind to MHC molecules with sufficient affinity to be presented to T cells. Currently, multiple tools such as SYFPEITHI, IEDB, and NetMHCpan are extremely useful in offering unique features in epitope-binding prediction. For instance, SYFPEITHI calculates the binding affinity of peptides to MHC class I molecules. It employs a scoring system based on experimental binding data to predict the likelihood of peptide-MHC interaction. This approach enables the identification of potential neoepitopes that have a high probability of being presented by MHC molecules and recognized by T cells. SYFPEITHI's approach, while valuable, is often constrained by the availability of experimental binding data for a diverse range of MHC alleles [18]. While SYFPEITHI, Rankpep, and BIMAS served as pioneering prediction tools, the field has seen the emergence of more refined alternatives. Among these, NetMHC stands out as one of the most widely utilized and rigorously validated algorithms available today. NetMHC employs artificial neural networks to predict peptide binding across various MHCI variants, yielding the predicted IC50 as an output. The accuracy of neural network-based methods relies on the training set's quality and size, thus performing better for more prevalent alleles. Notably, a refined version known as NetMHCpan expands the training dataset to encompass data from diverse species, enhancing the accuracy of predictions for less common MHC alleles [19].

Currently, the most useful epitope prediction algorithms are those focusing on peptide binding to MHC class I molecules. The MHCI antigen presentation pathway plays a central role in presenting peptides derived from endogenous cellular proteins to CD8+ T cells. Intracellular proteins undergo proteasomal processing, yielding 8–11 amino acid peptides that are subsequently transported into the endoplasmic reticulum (ER) by the transporter associated with antigen processing. There, they associate with newly synthesized class I molecules, forming stable peptide–MHCI complexes that are transported to the cell surface [20][21]. On the other hand, MHC class II antigen presentation involves the presentation of peptides derived from exogenous antigens, often proteins internalized through endocytosis or phagocytosis. In the endosomal compartments of antigen-presenting cells, these antigens are processed into peptide fragments, and a subset of these peptides binds to MHC II molecules within the groove created by the α and β chains. The resulting peptide-MHC II complex is then transported to the cell surface, where it is presented to CD4+ T helper cells. While prediction algorithms for MHCI neoantigens have flourished, MHC II neoantigens have posed challenges due to their diverse lengths (ranging from 13 to 25 amino acids) and increased binding complexity. As a result, there is a relative scarcity of binding-affinity training data and fewer algorithms available for predicting MHC II neoantigens [3].

5. Neoantigen Prioritization

Subsequent to neoantigen prediction using NetMHCpan, a pivotal step involves the comprehensive prioritization of neoepitopes, ensuring the selection of those with the highest potential to trigger a robust immune response. While NetMHCpan aids in identifying peptide sequences likely to bind to MHC molecules, further criteria are considered to assess their immunogenicity. Among the steps in neoepitope prioritization, the predicted binding affinity holds significance, as neoepitopes displaying strong MHC binding are more likely to be presented to immune cells. Tools like MHCflurry and MHCconsortium can also refine binding affinity predictions, aiding in the identification of top candidates. Additionally, assessing the conservation of the mutated amino acid across species using tools such as SIFT and PolyPhen enhances the understanding of its potential functional impact. Estimating epitope abundance is presently achieved through an indirect assessment involving the quantification of RNA expression levels. Mutations can be identified through tumor-to-normal DNA comparisons undergo bioinformatic scrutiny to gauge their immunogenic potential. The subsequent estimation of candidate immune stimulatory peptide levels is facilitated by RNA-Seq analysis. Prioritization based on gene expression levels, using databases like The Cancer Genome Atlas and Genotype-Tissue Expression, adds another layer of insight into neoepitope selection. Furthermore, considering the antigen processing machinery's efficiency, tools like NetChop and NetCTLpan evaluate proteasomal cleavage and T cell processing, respectively. Neoepitopes arising from frameshift mutations and non-synonymous alterations are often prioritized due to their potential to generate immunogenic peptides. Tailoring the prioritization based on tumor heterogeneity and patient-specific HLA type refines the strategy [22].

6. Neoantigen Validation

Validation of immunogenic neoantigens can be performed in many ways, among which the most common methodologies included are mass spectrometry, tetramer/multimer staining, and ELISpot, ELISA, or intracellular cytokine staining. By eluting bound peptides and identifying using tumor-specific variant libraries, mass spectrometry is able to profile the neoantigens presented on the MHC molecules. Its high sensitivity enables the detection of even minute quantities of antigens, making it well-suited for the task. Moreover, it offers an unbiased approach, capable of identifying a wide range of neoantigens without prior knowledge of their sequences. Mass spectrometry can also provide quantitative data about

the abundance of neoantigens, which is valuable for assessing their significance. However, it comes with certain complexities. Specialized equipment and expertise are prerequisites, making it less accessible for some laboratories. Additionally, sample preparation can be time-consuming and technically challenging. It is worth noting that mass spectrometry primarily detects neoantigens presented on MHC class I molecules, limiting its applicability to this subset of antigens [23].

Complementary to mass spectrometry, tetramer or multimer staining facilitates the visualization and quantification of neoantigen-specific T cells. This technique employs fluorescently labeled MHC–peptide complexes to detect and enumerate neoantigen-specific T cell populations, offering insights into their abundance and specificity and confirming the presence of T cells capable of recognizing the neoantigens. However, there are limitations to consider. Tetramer and multimer staining require prior knowledge of the neoantigens of interest and the availability of corresponding tetramers/multimers. Custom production of these reagents can be expensive and time-consuming. Moreover, these techniques may have limited sensitivity in detecting rare T cell populations, posing challenges in studies where low-frequency neoantigen-specific T cells are of interest [24].

Functional validation techniques like ELISpot, ELISA, and intracellular cytokine staining assess the ability of neoantigens to stimulate T cell responses. With synthesized neopeptides, ELISpot and ELISA quantify interferon-gamma secretion or cytokine production in response to neoantigens, providing quantitative data on T cell activation. Meanwhile, intracellular cytokine staining detects cytokine production within T cells, corroborating their activation status. However, they do have limitations to consider. ELISpot and ELISA may lack the specificity of tetramer staining and mass spectrometry, potentially leading to false-positive results. Intracellular cytokine staining, while capable of detecting functional responses, may not directly identify neoantigen-specific T cells. Additionally, the sensitivity of these techniques may be limited, particularly in detecting low-frequency neoantigen-specific T cell populations [25].

In conclusion, the choice of neoantigen validation technique should align with the specific goals of the study, available resources, and the nature of the neoantigens under investigation. Each technique has its strengths and limitations, and combining multiple methods can provide a comprehensive assessment of neoantigen-specific immune responses, enhancing the overall validation process.

References

1. Fritsch, E.F.; Burkhardt, U.E.; Hacohen, N.; Wu, C.J. Personal Neoantigen Cancer Vaccines: A Road Not Fully Paved. *Cancer Immunol. Res.* 2020, 8, 1465–1469.
2. Jiang, T.; Shi, T.; Zhang, H.; Hu, J.; Song, Y.; Wei, J.; Ren, S.; Zhou, C. Tumor neoantigens: From basic research to clinical applications. *J. Hematol. Oncol.* 2019, 12, 93.
3. Richters, M.M.; Xia, H.; Campbell, K.M.; Gillanders, W.E.; Griffith, O.L.; Griffith, M. Best practices for bioinformatic characterization of neoantigens for clinical utility. *Genome Med.* 2019, 11, 56.
4. Smith, C.C.; Selitsky, S.R.; Chai, S.; Armistead, P.M.; Vincent, B.G.; Serody, J.S. Alternative tumour-specific antigens. *Nat. Rev. Cancer* 2019, 19, 465–478.
5. Schumacher, T.N.; Scheper, W.; Kvistborg, P. Cancer Neoantigens. *Annu. Rev. Immunol.* 2019, 37, 173–200.
6. Biswas, N.; Chakrabarti, S.; Padul, V.; Jones, L.D.; Ashili, S. Designing neoantigen cancer vaccines, trials, and outcomes. *Front. Immunol.* 2023, 14, 1105420.
7. Carbone, D.P.; Reck, M.; Paz-Ares, L.; Creelan, B.; Horn, L.; Steins, M.; Felip, E.; van den Heuvel, M.M.; Ciuleanu, T.E.; Badin, F.; et al. First-Line Nivolumab in Stage IV or Recurrent Non-Small-Cell Lung Cancer. *N. Engl. J. Med.* 2017, 376, 2415–2426.
8. Lauss, M.; Donia, M.; Harbst, K.; Andersen, R.; Mitra, S.; Rosengren, F.; Salim, M.; Vallon-Christersson, J.; Torngren, T.; Kvist, A.; et al. Mutational and putative neoantigen load predict clinical benefit of adoptive T cell therapy in melanoma. *Nat. Commun.* 2017, 8, 1738.
9. Heemskerk, B.; Kvistborg, P.; Schumacher, T.N. The cancer antigenome. *EMBO J.* 2013, 32, 194–203.
10. Ho, S.Y.; Chang, C.M.; Liao, H.N.; Chou, W.H.; Guo, C.L.; Yen, Y.; Nakamura, Y.; Chang, W.C. Current Trends in Neoantigen-Based Cancer Vaccines. *Pharmaceutics* 2023, 16, 392.
11. Carreno, B.M.; Magrini, V.; Becker-Hapak, M.; Kaabinejadian, S.; Hundal, J.; Petti, A.A.; Ly, A.; Lie, W.R.; Hildebrand, W.H.; Mardis, E.R.; et al. Cancer immunotherapy. A dendritic cell vaccine increases the breadth and diversity of melanoma neoantigen-specific T cells. *Science* 2015, 348, 803–808.

12. Liu, X.S.; Mardis, E.R. Applications of Immunogenomics to Cancer. *Cell* 2017, 168, 600–612.
13. Thind, A.S.; Monga, I.; Thakur, P.K.; Kumari, P.; Dindhoria, K.; Krzak, M.; Ranson, M.; Ashford, B. Demystifying emerging bulk RNA-Seq applications: The application and utility of bioinformatic methodology. *Brief. Bioinform.* 2021, 22, bbab259.
14. De Mattos-Arruda, L.; Vazquez, M.; Finotello, F.; Lepore, R.; Porta, E.; Hundal, J.; Amengual-Rigo, P.; Ng, C.K.Y.; Valencia, A.; Carrillo, J.; et al. Neoantigen prediction and computational perspectives towards clinical benefit: Recommendations from the ESMO Precision Medicine Working Group. *Ann. Oncol.* 2020, 31, 978–990.
15. Cabanski, C.R.; Magrini, V.; Griffith, M.; Griffith, O.L.; McGrath, S.; Zhang, J.; Walker, J.; Ly, A.; Demeter, R.; Fulton, R.S.; et al. cDNA hybrid capture improves transcriptome analysis on low-input and archived samples. *J. Mol. Diagn.* 2014, 16, 440–451.
16. Kim, S.; Scheffler, K.; Halpern, A.L.; Bekritsky, M.A.; Noh, E.; Kallberg, M.; Chen, X.; Kim, Y.; Beyter, D.; Krusche, P.; et al. Strelka2: Fast and accurate calling of germline and somatic variants. *Nat. Methods* 2018, 15, 591–594.
17. Fan, Y.; Xi, L.; Hughes, D.S.; Zhang, J.; Zhang, J.; Futreal, P.A.; Wheeler, D.A.; Wang, W. MuSE: Accounting for tumor heterogeneity using a sample-specific error model improves sensitivity and specificity in mutation calling from sequencing data. *Genome Biol.* 2016, 17, 178.
18. Rammensee, H.; Bachmann, J.; Emmerich, N.P.; Bachor, O.A.; Stevanovic, S. SYFPEITHI: Database for MHC ligands and peptide motifs. *Immunogenetics* 1999, 50, 213–219.
19. Lundegaard, C.; Lamberth, K.; Harndahl, M.; Buus, S.; Lund, O.; Nielsen, M. NetMHC-3.0: Accurate web accessible predictions of human, mouse and monkey MHC class I affinities for peptides of length 8–11. *Nucleic Acids Res.* 2008, 36, W509–W512.
20. Van Kaer, L.; Ashton-Rickardt, P.G.; Ploegh, H.L.; Tonegawa, S. TAP1 mutant mice are deficient in antigen presentation, surface class I molecules, and CD4-8+ T cells. *Cell* 1992, 71, 1205–1214.
21. Robinson, J.; Halliwell, J.A.; Hayhurst, J.D.; Flicek, P.; Parham, P.; Marsh, S.G. The IPD and IMGT/HLA database: Allele variant databases. *Nucleic Acids Res.* 2015, 43, D423–D431.
22. Rammensee, H.G.; Singh-Jasuja, H. HLA ligandome tumor antigen discovery for personalized vaccine approach. *Expert Rev. Vaccines* 2013, 12, 1211–1217.
23. Zhang, X.; Qi, Y.; Zhang, Q.; Liu, W. Application of mass spectrometry-based MHC immunopeptidome profiling in neoantigen identification for tumor immunotherapy. *Biomed. Pharmacother.* 2019, 120, 109542.
24. Li, L.; Zhang, X.; Wang, X.; Kim, S.W.; Herndon, J.M.; Becker-Hapak, M.K.; Carreno, B.M.; Myers, N.B.; Sturmoski, M.A.; McLellan, M.D.; et al. Optimized polyepitope neoantigen DNA vaccines elicit neoantigen-specific immune responses in preclinical models and in clinical translation. *Genome Med.* 2021, 13, 56.
25. Yossef, R.; Tran, E.; Deniger, D.C.; Gros, A.; Pasetto, A.; Parkhurst, M.R.; Gartner, J.J.; Prickett, T.D.; Cafri, G.; Robbins, P.F.; et al. Enhanced detection of neoantigen-reactive T cells targeting unique and shared oncogenes for personalized cancer immunotherapy. *JCI Insight* 2018, 3, e122467.