Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Genetics & Development

Immunogenic peptide discovery in cancer genomes

Alexandra Snyder[1,2] and Timothy A Chan[2,3]

As immunotherapies to treat malignancy continue to diversify along with the tumor types amenable to treatment, it will become very important to predict which treatment is most likely to benefit a given patient. Tumor neoantigens, novel peptides resulting from somatic tumor mutations and recognized by the immune system as foreign, are likely to contribute significantly to the efficacy of immunotherapy. Multiple *in silico* methods have been developed to predict whether peptides, including tumor neoantigens, will be presented by the major histocompatibility complex (MHC) Class I or Class II, and interact with the T cell receptor (TCR). The methods for neoantigen prediction will be reviewed here, along with the most important examples of their use in the field of oncology.

Addresses
[1] Department of Medicine, Memorial Sloan Kettering Cancer Center, New York, NY, United States
[2] Human Oncology and Pathogenesis Program, Memorial Sloan Kettering Cancer Center, New York, NY, United States
[3] Department of Radiation Oncology, Memorial Sloan Kettering Cancer Center, New York, NY, United States

Corresponding author: Chan, Timothy A (chant@mskcc.org)

Current Opinion in Genetics & Development 2015, **30**:7–16

This review comes from a themed issue on **Cancer genomics**

Edited by **Christine A Iacobuzio-Donahue** and **Elaine A Ostrander**

For a complete overview see the Issue

Available online 12th January 2015

http://dx.doi.org/10.1016/j.gde.2014.12.003

Introduction: why prediction, why now?

When William B. Coley, the 'father' of immunotherapy, injected streptococcal organisms into patients with metastatic solid tumors in the 1890s, it was not known what aspects of 'Coley's Toxins,' the patient and/or tumor, were responsible for tumor regression in a subset of patients with metastatic cancer. Nearly a century later, IL-2 was approved for the treatment of metastatic melanoma [1]. Since then, checkpoint blockade therapies and T cell therapies have proliferated, with promising results for both [2,3,4••]. With this increase in immunotherapies, clinicians will need tools to predict which type of immunotherapy is most likely to benefit a specific patient.

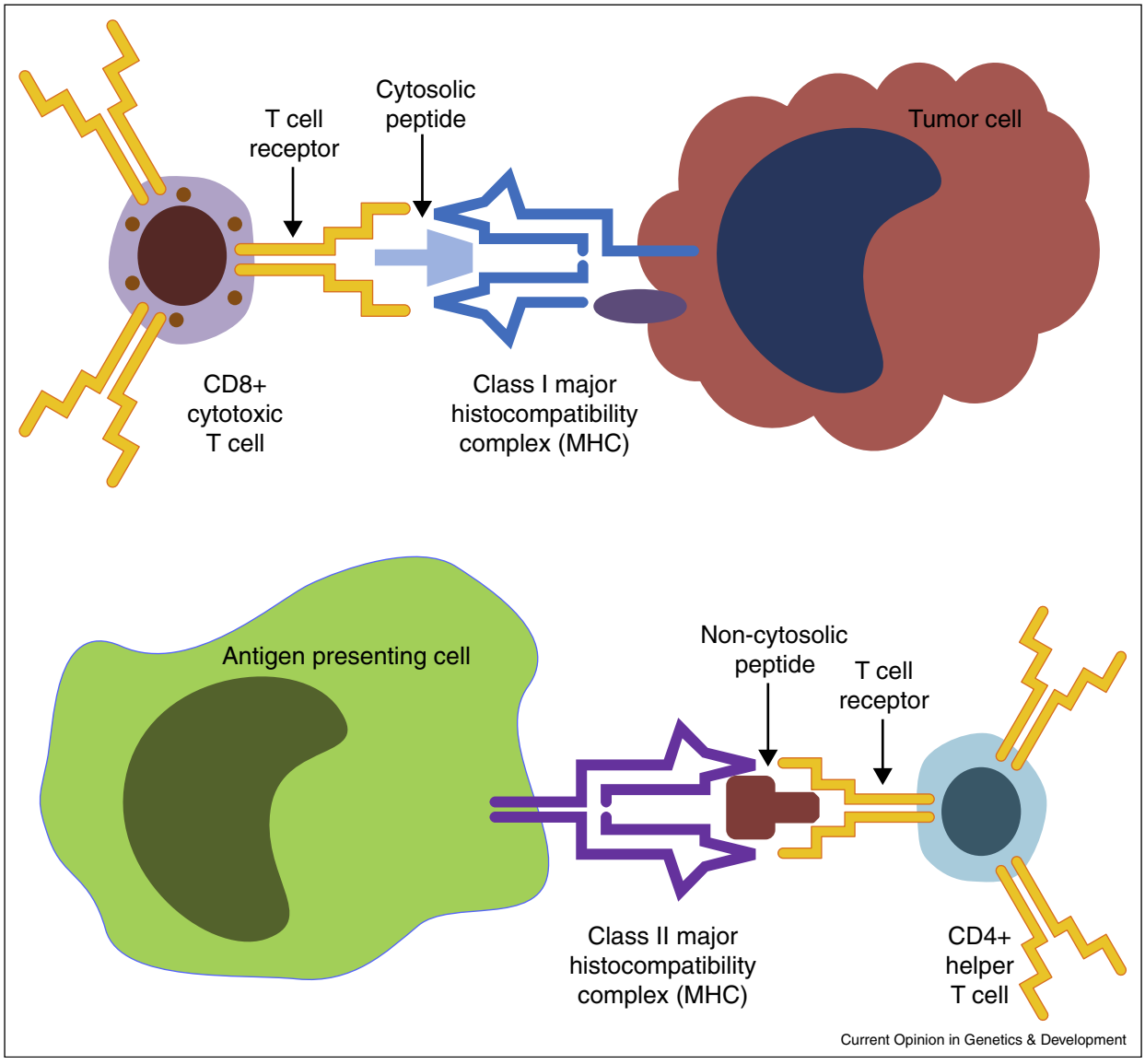
A growing body of literature suggests that response to multiple types of immunotherapy results from the anti-tumor immune response against a critical neoantigen(s), non-self peptides resulting from exonic missense mutations (reviewed in [5•]). Cancer vaccines have historically used tumor-associated antigens which are overexpressed in tumors and have restricted tissue expression. However, these therapies require overcoming central and peripheral tolerance. In contrast, a tumor neoantigen would in theory not be limited by tolerance, with data both in preclinical models and humans supporting this idea [6•,7•]. It is important to note that not all effective immunotherapies target tumor neoantigens, but rather may target viral [8] or other antigens, as with chimeric antigen receptors [9,10] and therapeutic antibodies [11].

The major hurdle to selecting, improving and designing immunotherapies based on neoantigens lies in their accurate prediction, a challenging process in light of the complexities of the immune system. The major histocompatibility complex (MHC) molecules include two classes. MHC Class I (in humans, human leukocyte antigen, or HLA, Class I) molecules bind intracellular antigens of 8–11 amino acids in length and present them to cytotoxic CD8+ T cells. MHC Class II molecules bind extracellular antigens of 11–20 amino acids and present them to T helper CD4+ cells [12]. The HLA alleles are incredibly diverse, with greater than 6000 HLA Class I and Class II alleles described to date [13]. The number of potential peptides processed from a given pathogen is also vast, with a small proportion actually binding MHC Class I or II [14]. These facts make peptide prediction both important and challenging.

An immunogenic peptide fulfills at least two criteria: presentation by an MHC molecule and recognition by a T-cell receptor (Figure 1). In order to be presented by the MHC, a protein may be cleaved, and is typically presented by the antigen presentation machinery. Multiple computational algorithms exist for each step in this process. A fairly exhaustive list of prediction tools is provided (http://cancerimmunity.org/resources/webtools/ and tables in [15–17]), and the bioinformatic aspects of many of these programs has been reviewed [17,18•,19•].

In the absence of data showing which peptides are presented by a given MHC, whole exome sequencing can be used to indirectly predict this. The sequenced DNA can be 'virtually translated' into predicted proteins, then the analytic tools described here evaluate which peptides may be presented by a given patient's HLA repertoire (Figure 2). Here, we will first describe and evaluate the most commonly used MHC Class I, Class II and T cell predictors, highlighting those with data to support their

Figure 1



Summary of antigen presentation. A tumor cell (top) loads a peptide onto the Class I major histocompatibility complex (MHC) for presentation to and interaction with a CD8+ cytotoxic T cell. An antigen presenting cell (APC) loads a peptide onto the Class II MHC for presentation to and interaction with a CD4+ helper T cell (right). Additional co-receptor interactions between the APC and each type of T cell are not displayed here. These interactions play an important role in determining the downstream fate of each T cell, but as yet cannot be predicted bioinformatically.Image based on figures from motifolio.com.

use both in infectious diseases and oncology. We will then describe the most important recent uses of these tools in preclinical and clinical oncologic settings.

MHC Class I prediction

There are many MHC Class I prediction tools (Table 1). Most have been trained on data from the Immune Epitope Database (IEDB) [20]. There are two general categories of prediction tools: allele-specific and pan-specific. Initial programs relied on allele-specific motifs; for example, positions 2 and 9 constitute anchor residues on HLA-A*02:01, commonly occupied by leucine, valine or isoleucine [21]. Other positions are similarly stereotyped [22,23] and unknown input peptides searched for allele-specific motifs. A common problem with these methods is that there are insufficient data from rare alleles to reliably predict peptides which will bind to them. Three-dimensional structural models have been designed, but thus far underperform the models trained on actual data [19•].

To address this problem, pan-specific programs were created to extrapolate from existing data to less common