Front. Biol. 2016, 11(5): 366–375
DOI 10.1007/s11515-016-1422-2

REVIEW

Finding neoepitopes in mouse models of personalized cancer immunotherapy

Sahar Al Seesi[1,2,*], Alok Das Mohapatra[1,*], Arpita Pawashe[1], Ion I. Mandoiu[2], Fei Duan (✉)[1]

[1] *Department of Immunology and Carole and Ray Neag Comprehensive Cancer Center, University of Connecticut Cancer Center, Farmington, CT 06030, USA*
[2] *Department of Computer Science & Engineering, University of Connecticut, Storrs, CT 06269, USA*

BACKGROUND: Cancer immunotherapy uses one's own immune system to fight cancerous cells. As immune system is hard-wired to distinguish self and non-self, cancer immunotherapy is predicted to target cancerous cells specifically, therefore is less toxic than chemotherapy and radiation therapy, two major treatments for cancer. Cancer immunologists have spent decades to search for the specific targets in cancerous cells.
METHODS: Due to the recent advances in high throughput sequencing and bioinformatics, evidence has merged that the neoantigens in cancerous cells are probably the cancer-specific targets that lead to the destruction of cancer. We will review the transplantable murine tumor models for cancer immunotherapy and the bioinformatics tools used to navigate mouse genome to identify tumor-rejecting neoantigens.
RESULTS: Several groups have independently identified point mutations that can be recognized by T cells of host immune system. It is consistent with the note that the formation of peptide-MHC I-TCR complex is critical to activate T cells. Both anchor residue and TCR-facing residue mutations have been reported. While TCR-facing residue mutations may directly activate specific T cells, anchor residue mutations improve the binding of peptides to MHC I molecules, which increases the presentation of peptides and the T cell activation indirectly.
CONCLUSIONS: Our work indicates that the affinity of neoepitopes for MHC I is not a predictor for anti-tumor immune responses in mice. Instead differential agretopic index (DAI), the numerical difference of epitope-MHC I affinities between the mutated and un-mutated sequences is a significant predictor. A similar bioinformatics pipeline has been developed to generate personalized vaccines to treat human ovarian cancer in a Phase I clinical trial.

Keywords cancer immunotherapy, tumor antigens, neoantigens, neoepitopes, differential agretopic index (DAI), RNA-Seq, single nucleotide variant (SNV)

Introduction

Cancer is a generic term used for a large group of diseases defined by uncontrolled division, multiplication and metastasis of transformed cells, owing to accumulation of multiple intracellular genetic aberrations. Unraveling these genetic aberrations and understanding their contribution to the onset and progression of cancer is considered crucial for augmenting existing approaches to diagnosis, treatment and prevention of cancer. Another alternative avenue to realize the same objective is to appreciate the intricate potential of host immune system to recognize and eliminate cancer. Animal models have now become indispensable to study human diseases including cancer not only to overcome the moral and ethical obstacles associated with human experimentation but also to address the issue of toxic side effects of emerging anti-cancer therapies (Dranoff, 2012). A number of animal models have been developed to study cancer. They include invertebrate animal models like worms and flies and vertebrate animal models including rodents (such as mouse), dogs, chickens and zebra fishes. However, more than any other model systems mice have revolutionized our

Received April 25, 2016; accepted August 15, 2016
Correspondence: Fei Duan
E-mail: fduan@uchc.edu

[*]These authors contributed equally to this work.

ability to understand the function of oncogenes in vivo and their precise role in pathogenesis of cancer. They have also provided critical insights into the cellular and molecular mechanisms underlying immune rejection of cancer, which has led to emergence of a promising branch of therapeutic intervention in cancer, commonly known as cancer immunotherapy. As a model system, mice offer several advantages over other mammalian models (Cheon and Orsulic, 2011; Berman et al., 2014) because of their (a) smaller size, (b) easy and economic maintenance, (c) fairly reasonable breeding capacity and large litter size, (d) acquiescence to genetic manipulation, (e) suitability for xenograft studies, (f) availability of viral and spontaneous tumor models and (g) extensive genetic similarity to humans.

In the early part of twentieth century, the studies attempted to graft human cancers in large animals were unsuccessful due to naturally occurring xenogeneic barriers across different species. However, when T lymphocytes were later identified as the cellular effectors of xenorejection, experimentally induced T lymphopenia by thymectomy was used to generate immunodeficient mice for human cancer xenografts in conjunction with depletion of mature lymphocytes (using anti lymphocyte globulin). In the 1960s, spontaneous immune-deficient strains of mice were identified and became the natural successors as xenograft hosts. The T lymphocyte deficient nude mice identified in 1962 allowed successful and large-scale transplantation of human tumors and cancer cell lines into mouse hosts. Discovery of more strains of mice that are immune-deficient (e.g. Nude/SCID, NOD/SCID) including those which harbor either spontaneous mutations (e.g. *Prkdc*^{scid}, *beige*) or targeted mutations (e.g. *rag1*^{null}, *rag2*^{null}) have provided enhanced opportunity for xenotransplantation of a variety of human hematological and solid tumors in mouse models (Berman et al., 2014).

In parallel, non-inbred rodents were replaced by inbred mice strains in the 1950s, to differentiate specific tumor rejection from allogeneic tissue rejection (Blanchard et al., 2013). The initial demonstration of rejection of chemically induced sarcomas (using methylcholanthrene as a chemical carcinogen) in syngeneic mice was published in 1953 (Foley, 1953). Careful studies done later using chemically induced tumors in mouse models revealed the specificity of antitumor immunity and ushered in a new hope to fight cancer by exploiting the arsenals of host immune system (Prehn and Main, 1957; Basombrio,1970). In the mid-1950s, genetic basis of cancer was demonstrated with the discovery of oncogenic viruses in mice, which made them increasingly susceptible to cancer. Later on, the identification of tumor suppressor and tumor promoting genes as well as development of transgenic and gene targeting technologies in mouse embryonic stem cells in the 1980s facilitated the generation of genetically engineered mouse (GEM) models to study tumor biology through the manipulation of the mouse germ line (Cheon and Orsulic, 2011).

The proof of concept studies for new cancer therapies arise, mostly from mechanical insights obtained from experimental studies performed in mouse models. However, availability of a wide array of mouse models raises several important questions (Dranoff, 2012). How to choose which systems to study? Whether any system faithfully recapitulates tumor development in human subjects? Can the lessons learned from one mouse model be extrapolated to other mouse models? How informative comparative studies of different therapies can be in the same model? Given the growing number of potential targets and strategies for therapeutic intervention in cancer, these difficult questions need some urgent attention from the scientific community. In the present review we have attempted to focus on the most commonly used experimental mouse model: transplantable mouse tumor models.

Transplantable mouse models of cancer

These mouse models have been in the forefront of cancer research for several decades and support tumor growth in them when inoculated with single cell suspensions of a given tumor cell line or when transplanted with a small piece of tumor. Single cell suspension of tumor cells is usually inoculated subcutaneously, intradermally or intravenously. Recent advances in imaging technology, which enable monitoring of tumor growth *in vivo*, permit orthotropic injections of tumor cells (injection of tumor cells in to organs of tumor origin, e.g. intracranial injection of glioma cells or intrarenal injections Renca cells) in these models (Dranoff, 2012; Liu et al., 2014). This approach allows growth of tumors in their natural microenvironment and hence is more physiologically relevant. The most common advantages and disadvantages of transplantable mouse tumor models are listed in Table 1.

Specificity in recognition and elimination of target cells or molecules by immune system has been successfully exploited by immunologists to devise effective vaccines against a range of infectious agents including cancers of infectious (viral) etiology (Srivastava, 2015). However, this issue of specificity has emerged as a challenging problem in case of human cancers, most of which have non-infectious etiology. Use of syngeneic inbred mice and transplantable tumor models revealed, for the first time that mice could be immunized against syngeneic tumors or could be used to demonstrate generation of specific anti-tumor immunity. This phenomenon was found to hold true for carcinogen (e.g. methylcholanthrene, UV, oncogenic viruses) induced tumors. Spontaneous tumors also exhibit specific immunity though to a lesser extent.

The success gained in immunization studies in experimental mice models, using autologous irradiated cancer cells and heat shock protein (HSP)-peptide complexes (harvested

Table 1 The advantages and disadvantages of transplantable murine tumor models

Advantages of transplantable murine tumor model	Disadvantages of transplantable murine tumor models
Transplantable mouse tumor models allow rapid and reliable tumor growth.	Rapid growth of tumor might prevent normal interaction between tumor cells and immune system.
They can be efficiently used to evaluate in both prophylactic and therapeutic models (in terms of altered tumor growth).	Tissue damage associated with injection and death of tumor cells might induce inflammation and alter immune response to tumors.
Induced expression of model antigens (such as ovalbumin) or viral gene products in tumor cell lines permit better understanding of anti-tumor immune response.	The intrinsic immunogenicity and tumor microenvironment of transplantable tumor cell lines shows wide variation from poorly immunogenic (e.g. B16 melanomas), moderately immunogenic (EL4 T lymphomas) and immunogenic (methylcholanthrene induced fibrosarcomas).
Models for various cancer types (e.g. melanoma, lymphoma, sarcoma, colon carcinoma etc.) are available.	They serve as a weaker model of natural tumor microenvironment. (This limitation can be overcome to some extent by orthotropic injections).

from autologous cancer cells) created much excitement among scientific community in the 1980s and 1990s. This was logically followed by several clinical trials based on the above two approaches, but yielded mixed outcomes. However, the original question about specificity in human cancers still remained unanswered (Srivastava, 2015). In the mid-1990s, research conducted in transplantable mouse tumor models lead to proposition of two common mechanisms: mutations and abnormal transcription, which are otherwise suppressed in normal tissues, to generate tumor specific antigens (Boon and van der Bruggen, 1996). Subsequent work in mouse models led to identification of different types of tumor antigens for immunotherapeutic intervention.

Tumor antigens

Shared cancer antigens

These antigens are so named as they are either shared among different cancers or between cancerous and normal cells. These antigens may be derived from lineage specific differentiation antigens (such as tyrosinase, Melan-A/Mart-1, gp100, gp75 and TRP-2) or from cancer testis antigens including 14 families and 53 cancer testis genes as of 2004 (such as MAGE-A, MAGE-C1, BAGE, GAGE, CAGE, XAGE, NY-ESO1/LAGE1) (Blanchard et al., 2013). These genes are frequently expressed in a wide variety of tumor types and also in some normal tissues including testicular germ cells, placental trophoblasts and medullary thymic epithelial cells (Boon and van der Bruggen, 1996; Srivastava, 2015). One vaccine using MAGE-A3 has been tested in two large Phase III trials, one for melanoma (ClinicalTrials.gov, Identifier: NCT00796445) and the other one for non-small cell lung cancer (NSCLC, ClinicalTrials.gov, Identifier: NCT00480025), *albeit* the risk of breaking self-tolerance and commencement of auto immunity. Unfortunately both failed to extend disease-free survival significantly when compared with placebo. Another vaccine targeting NY-ESO-1 also failed to improve relapse free survival in a Phase II trial for melanoma (ClinicalTrials.gov, Identifier: NCT00199901).

Cancer neoantigens

These antigens owe their origin to point mutations, translocations, insertions or deletions in individual cells of a tumor and are entirely specific for that tumor. Mutated self-polypeptides have been reported as specific tumor-rejection antigens in an anecdotal manner (Lurquin et al., 1989; Monach et al., 1995). Unfortunately we know remarkably little about spontaneous T cell responses against mutations. To characterize natural T cell responses to mutated polypeptides, Duan et al. (2009) mutagenized tyrosinase related protein 1(Tyrp1), an abundant and functionally important self-antigen expressed by melanocytes. The mutant construct was designated *White Magic* (WM) because it broke tolerance to self-antigen, Tyrp1, and induced robust T cell responses against melanocytes and destroyed them. As a result, the coat color of immunized C57BL/6 mice turned into white, a phenomenon called *vitiligo*, a known tolerable side effect when the immunotherapy against melanoma is effective. Two mouse tumor cell lines were transfected with WM for the study. One was spontaneously arising B16 melanoma and the other one was an immunogenic, chemically induced LiHa fibrosarcoma. Two major findings were: 1) LiHa-WM elicited strong spontaneous T cell responses, which lead to tumor rejection; $CD4^+$ and $CD8^+$ T cells were required for this process. 2) B16-WM by itself failed to induce either effective T cell responses or any tumor regression. With extra help by targeting regulatory T cells, B16-WM induced both T cell responses and tumor rejection, suggesting the impact of tumor microenvironment on spontaneous T cell responses against mutated polypeptides.

Similar results from immunization studies in mouse models with tumor specific epitopes have been quite encouraging. Although these results corroborate the proof of principle for immunotherapeutic intervention in cancer, due to their extremely limited scope of application (which is confined to individual tumors harboring such unique mutations), they did not elicit much enthusiasm among the scientific community (Srivastava, 2015) until recently.

In 1993, Dr. Pramod Srivastava, one of the pioneers in cancer immunology, stated the first time that “the specificity

of immunogenicity of tumors resulting from different random mutations in each cell" (Srivastava, 2015). According to his hypothesis, lack of complete fidelity in the DNA repair and replication mechanisms permits accumulation of thousands of random passenger mutations in the constituting cells of a growing tumor. Some of these mutations can indiscriminately generate new epitopes. Because of the randomness of this phenomenon, each new tumor can harbor a unique set of mutations as well as epitopes or neoepitopes. With the advent of high throughput sequencing and bioinformatics algorithms, it is now possible to test this hypothesis. Recently, several studies have addressed this problem with the same basic approach with little modifications (Castle et al., 2012; Matsushita et al., 2012; Duan et al., 2014; Gubin et al., 2014; Yadav et al., 2014). Most of these studies were performed in transplantable mouse models of cancer (e.g. methylcholathrene induced fibrosarcoma).

The common steps include: 1) identification of single nucleotide variants (SNVs) by comparing the expressed sequences of tumors to the normal genomes; 2) prediction of neoepitopes from these SNVs using algorithms that predict binding affinity of peptides to the major histocompatibility complex (MHC) molecules; and 3) testing the immunogenicity of these peptides in terms of their ability to elicit $CD8^+$ T cell response as well as tumor rejection in mice. The focus of our group was on defining the rules to predict tumor-rejecting epitopes recognized by T cells from cancer genome. The major findings from this project were: the predicted affinity of neoepitopes for MHC I had no bearing on protective antitumor immunogenicity in mice; instead, the numerical difference of such affinities between the mutated and unmutated sequences, designated *Differential Agretopic Index* (DAI), was a significant predictor (Duan et al., 2014). One surprising finding was that the epitopes bound to MHC I molecules with high affinity, or at low IC_{50}, do not provide tumor protection. Instead, the epitopes bound to MHC I molecules with medium affinity provide tumor protection. It has been demonstrated in a T cell receptor (TCR) transgenic model that $CD8^+$ T cell activation is governed by TCR-pMHC (peptide-MHC complexes) affinity (Tian et al., 2007). It is possible that the peptides bound to MHC I molecules with medium affinity engage high affinity TCRs while the peptides bound to MHC I molecules with high affinity engage low affinity TCRs. This hypothesis can be tested by isolating TCRs that recognize very stable pMHC complexes versus "not-so-stable" pMHC complexes and measure the binding of TCRs to pMHC complexes.

To reconcile recent studies on T cell recognition of neoepitopes, we have revealed that the dissimilarity of a neoepitope from germline, termed DAI, is an effective, yet imperfect, predictor for mutated-self cancer neoepitopes. The neoepitopes that we have identified were derived from anchor residue mutations that improve the binding of neoepitopes to MHC I molecules. This may be due to our bias in using NetMHC, one component of our bioinformatics pipeline. Gubin et al. (2014) have identified two neoepitopes, one of which is derived from a mutation in TCR-facing residues. Although the mutations are at epitope position P1 and P4, not the anchor positions (P5 and P8 for H-$2K^b$ molecule), they did deprioritize those epitopes that displayed lower binding affinity to class I than their corresponding wild-type sequences. Yadav et al. (2014) have used mass spectrometry as a filter for the neoepitopes actually presented by the tumor. They identify three neoepitopes. Two neoepitopes have mutations that generate TCR-facing residues, and one generates an optimized anchor residue.

Notably, Yadav et al. (2014) and Gubin et al. (2014) focus exclusively on peptides meeting a specific affinity cutoff, while our study has tested, unbiasedly, the protective immunity that may be generated by potential neoepitopes of both high and low predicted MHC binding affinities. The prediction algorithms developed by all three groups are imperfect, and all need very significant refinement in bioinformatics, MHC I-peptide interaction, MHC I-peptide-TCR interaction, and immunity and tolerance.

Heterogeneity of neoepitopes

Not all mutations are created equal. So are neoepitopes derived from non-synonymous mutations. Cancer cells contain numerous random mutations (Bielas et al., 2006) as well as clonal mutations (Nowell, 1976), or branch versus trunk mutations from a phylogenetic tree of cancer (Yates and Campbell, 2012). Clonal mutations are present in most of or all tumor cells and assumed to be selected to confer a growth advantage. Clonal driver mutations are appealing targets for therapy because they often shared by different tumors during oncogenesis. But one can also argue that an evolved tumor may have already acquired the resistance to T cells recognizing those clonal mutations. For example, a point mutation in p53 at amino acid position 234 (M234I) generate a neoeptiope (KY**I**CNSSCM) in Meth A fibrosarcoma on BALB/c mouse background (Noguchi et al., 1994). But immunization against Meth A tumor only caused 7-day delay of tumor growth not rejection.

In a mouse fibrosarcoma model, 30 distinct clones were derived from mouse Meth A fibrosarcoma line (Duan et al., 2014). We sequenced the genomic DNA samples from all the clones to confirm four mutations identified by RNA-Seq. We found 29/30 clones harbored 4/4 mutations and only 1/30 clone contained 3/4 mutations, strongly suggesting the clonal mutations induce tumor immunity. This finding in mouse model is supported by a recent study in non-small cell lung cancer (NSCLC) and melanoma in which clonal neoantigen burden correlates with clinical benefit and sensitivity to immune checkpoint blockade (anti-PD-1 and anti-CTLA-4) (McGranahan et al., 2016).

Our thoughts on this translatable finding are twofold. 1) The lack of antigenic heterogeneity is due to the relatively shallow depth of sequencing. 2) Most importantly, these

results suggest it is possible to use a relatively shallow sequencing as a methodology to identify the neoepitopes that are the most broadly distributed among cancer cells.

The random mutations, or unique mutations within a tumor however, may actually delay tumor growth, for instance, by eliciting senescence, strong immune responses and bystander killing. There is evidence that neoepitopes are recognized by T cells, which drives the immunological destruction or sculpting of a developing cancer (Matsushita et al., 2012).

Immune checkpoint blockade in cancer therapy

The current enthusiasm surrounding cancer immunotherapy stems from the unprecedented clinical success of immune checkpoint receptor blockade (e.g. CTLA-4 blockade) against a range of tumors. The proof of concept studies were initially obtained using moderately immunogenic transplantable cell lines such as TRAMPC1 (also known as pTC1) prostate carcinoma, ovalbumin expressing lymphoma and Sa1N fibrosarcoma cells. Similarly, blockade of another immune checkpoint receptor (e.g. PD1 blockade) have also been shown to have promising anti-tumor effects in varied human cancers including melanoma, lymphoma, colon carcinoma and renal carcinoma. Astonishingly, in case of PD1 blockade the patient benefits exceeded the expectations compared to the findings from transplantable mouse models (Dranoff, 2012). Gubin et al. (2014) provided first evidence in mouse sarcoma models that anti-PD-1 and/or anti-CTLA-4 therapy is effective because neoepitopes serve as a major class of T cell rejection antigens. This has been confirmed in patients with non-small cell lung cancer (NSCLC) and melanoma who responded to immune checkpoint blockade (anti-PD-1 and anti-CTLA-4) (McGranahan et al., 2016). Overall, the observed concordance between findings from pre-clinical mouse models and patients in checkpoint blockade studies was quite encouraging, suggesting that transplantable tumor models can provide important clues for successful clinical translation. In contrast, the checkpoint blockade using CTLA-4-specific antibodies exhibited poor outcomes in a less immunogenic B16 melanoma model (Dranoff, 2012). Hence, a better understanding of the differential immune response between poorly immunogenic and immunogenic tumor models might provide essential clues to find novel approaches to increase patient survival.

Bioinformatics analysis

Next generation sequencing made it possible to sequence the complete genome and transcriptome at single base resolution, revolutionizing the cancer genomics and immunotherapy research. There are multiple sequencing platforms available, such as SOLiD (Pandey et al., 2008), 454 (Thomas et al., 2006), Illumina (Bentley et al., 2008), and ION Torrent (Boland et al., 2013). Continuous technological advances that lead to longer reads, higher sequencing depth and lower cost, are making next generation sequencing the method of choice in studying cancer genomics.

Next generation sequencing data can be used in addressing multiple questions in the cancer genomics and immunotherapy research, including finding driver mutations, studying gene expression profiles in tumors, and identifying somatic variants that give rise to cancer specific epitopes. There are many different sequencing protocols addressing specific questions. Here we will focus on whole genome, exome, and transcriptome sequencing (RNA-Seq).

Whole genome sequencing can be used to call single point and structural variants, determine copy number variations in tumor samples, and other applications. If variants in regulatory regions are of no interest, genome sequencing can be replaced by exome sequencing. The cost of exome sequencing is much less since the size of the exome is about 1% of the genome. RNA-Seq is used to study the expression levels of genes as well as identify genes with tumor-specific expression patterns. In mouse models, RNA-Seq can also be used to call somatic variants without the need for genomic data. We will focus on bioinformatics analysis for identification of tumor specific epitopes.

Read quality control

Analysis pipelines usually start with quality control on sequencing reads. This step removes low quality bases that can appear at either end of the reads. The whole read can also be filtered out, as a low quality read, if the read has many low quality bases. Identical reads, indicating the bias in the library preparation step; should also be filtered. Figure 1 is a cartoon depiction of the read quality control step. Some of the available tools for read quality control are FASTX (http://hannonlab.cshl.edu/fastx_toolkit/index.html), PRINSEQ (Schmieder and Edwards, 2011) and Trimmomatic (Bolger et al., 2014).

Read mapping

A very important step in any analysis pipelines is the read mapping step. The choice of mapping software, reference and mapping parameters depends on the type of the sequencing data (transcriptome versus genomic), as well as the sequencing technology. The problem being addressed can also affect some of these decisions. Whole genome must be mapped unspliced to a genome reference, using an aligner like Bowtie (Langmead et al., 2009). On the other hand, trying to map RNA-Seq using unspliced aligner to the genome will result in losing all reads coming from exon-exon junction regions in the transcripts. Therefore, it can be mapped to transcriptome reference (a reference containing the gene/transcript

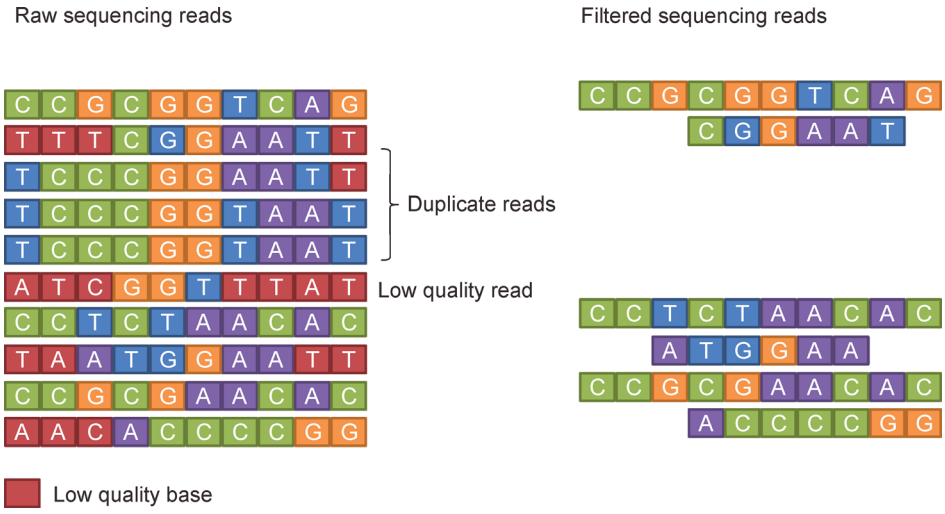

Figure 1 Read quality control. Raw sequencing reads are examined for low quality bases and low quality reads. Low quality bases at the reads' ends are trimmed. Reads containing a large percentage of low quality bases are discarded. Duplicate reads are also discarded. Duplicate reads are the results of sequencing protocol bias and may lead to false positive calls.

sequences with the introns spliced out). Alternatively, RNA-Seq reads can be mapped to the genome using a spliced aligner like TopHat (Trapnell et al., 2009) or HISAT2 (Kim et al., 2015), or an aligner that support local alignments like TMAP (https://github.com/iontorrent/TS/tree/master/Analysis/TMAP). Exome sequencing can be mapped unspliced to both genome and transcriptome sequences. Mapping the transcriptome is naturally much faster due to the small size of the transcriptome, compared to the genome. The type of sequencing errors of the technology being used affects the choice of mapper. For example, ION Torrent sequencers are known to have insertion and deletion (indel) errors in their reads, which makes TMAP a more suitable aligner for this type of data, compared to Bowtie or TopHat. Finally, the problem being addressed also affects the choice of mapping settings. For predicting SNVs, only reads that map to unique positions in the genome are considered in order to improve the confidence in the calls. Also, a higher rate of mismatch is usually allowed and these mismatches are considered SNV candidates. When RNA-Seq reads are used to estimate gene expression levels, mapping should be more stringent in terms of the number of allowed mismatches. Discarding non-uniquely mapped reads can negatively affect the accuracy of the gene expression estimates. Therefore, state-of-the-art gene expression prediction methods, like RSEM (Li and Dewey, 2011) and isoEM (Nicolae et al., 2011) are capable of handling the uncertainties in non-uniquely mapped read. Figure 2 illustrates a decision tree with different factors that affect read mapping decisions.

Gene expression analysis

Tumor transcriptome profiling is used to identify genes whose expression is associated with a certain type of tumor. There are multiple tools that perform gene expression quantification of RNA-Seq samples (Li and Dewey, 2011; Nicolae et al., 2011; Roberts et al., 2011; Kim et al., 2015). Gene differential expression (DE) analysis can be used to compare the transcriptome profiles of tumor samples against control samples, or to study the transcriptome profile changes across different time points. Some of the available DE tools include IsoDE (Al Seesi eet al., 2014), edgeR (Robinson et al., 2010), and GFOLD (Feng et al., 2012).

Calling somatic variants

Cancer immunotherapy research involves predicting immunogenic epitopes at somatic variants loci. Inbred mouse strains with known genome sequences are used in mouse cancer research. Therefore, calling somatic variants can be done by comparing the tumor sequencing data with the reference genome sequence, unlike human cancer research, where differentiation between somatic variants and germline mutations requires the joint analysis of sequencing data from tumors and matched normal tissue. Furthermore, the ability to call variants by comparison to the reference genome makes it possible to identify somatic variants in mouse tumor models using RNA-Seq rather than DNA sequencing data. Since tumors and normal tissues express different genes, it is unfeasible to distinguish somatic and germline variants from matched tumor-normal RNA-Seq samples.

GATK (McKenna et al., 2010) is a variant calling tool capable of calling SNVs and indels from genomic sequencing data. Calling variants from transcriptomic data are a more challenging problem due to the non-uniform read coverage (coverage will depend on gene expression). SNVQ (Duitama et al., 2012) is another variant calling tool. It calls only SNVs but works equally well with genomic and transcriptomic data.

For cancer immunotherapy, the aim is to identify variants that will result in presenting tumor specific epitopes on tumor cells (neo-epitopes). These variants must be in expressed genes, and precisely the set of variants that appear in the RNA-Seq data. If genomic (whole genome or exome) data was used to predict variants, then RNA-Seq data must be used to check expression of genes bearing these variants.

Epitope prediction and vaccine design

This step starts by predicting a pool of candidate epitopes that can be used in vaccine design. This is done through applying the identified somatic variants to proteins then passing these modified protein sequences to epitope prediction tools. Tools like SYFPEITH (Schuler et al., 2007) and NetMHC (Lundegaard et al., 2008) will generate an IC_{50} score that indicates how strongly the peptide would bind to MHC molecules, which is a requirement for the peptide to be presented on the tumor cell surface. NetCTL (Larsen et al., 2007) is another epitope prediction tool that also takes into account the likelihood of the peptide to be generated inside the cell upon protein being cleaved by the proteasome.

The epitope prediction tools generate a large list of epitopes. Selecting which epitopes would be the best to go into the vaccine is a challenging task. The goal is to maximize the efficacy of the vaccine. Duan et al. (2014) showed that selecting epitopes with the highest difference in MHC binding score between the mutant epitope and its wild type counterpart (differential agretopic index or DAI) increased the probability of the epitope to have tumor rejection

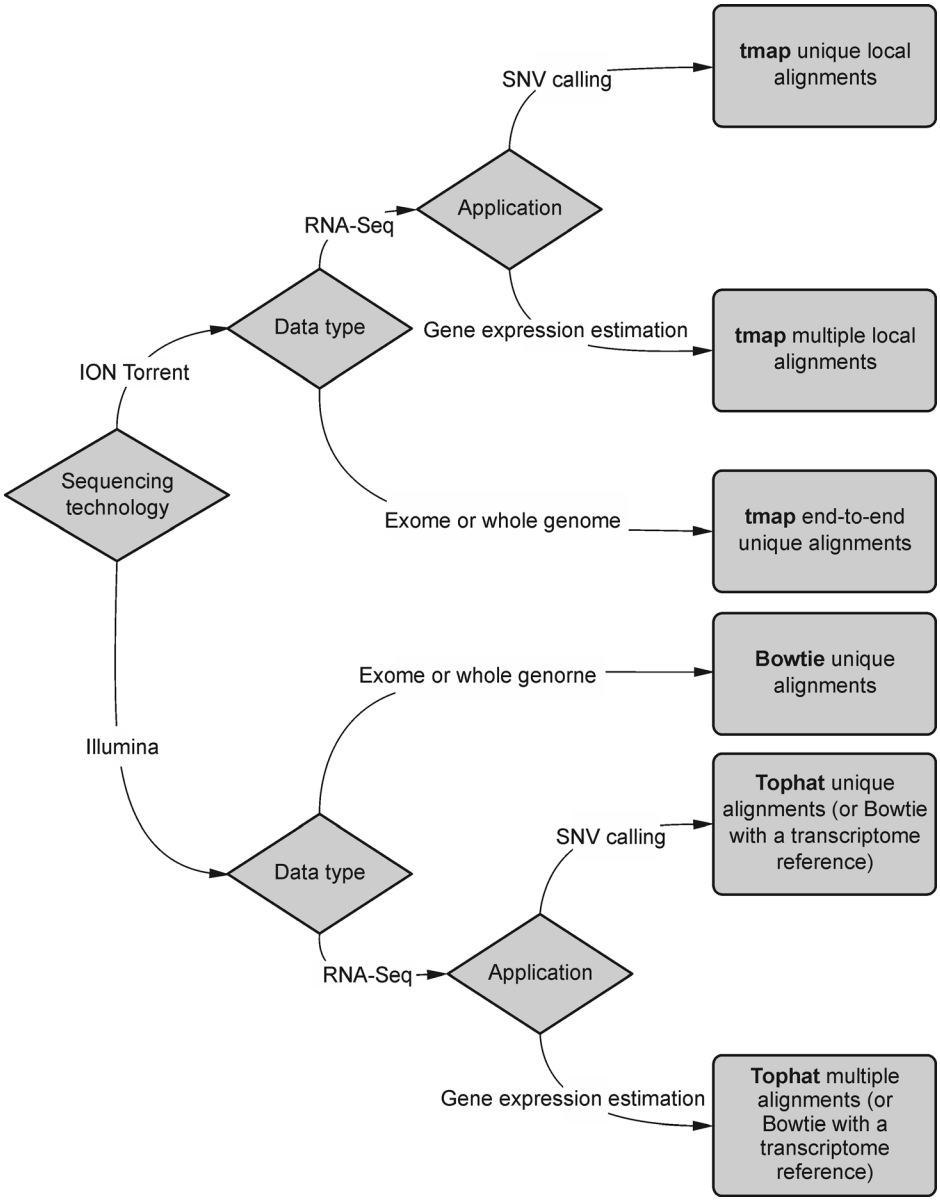

Figure 2 Mapping decision tree. This figure illustrates the effect of different factors on choosing a read aligner and setting the mapping parameters. Factors include the sequencing technology and platform, as well as the question being studied.

response. Since not all tumor cells will harbor all candidate epitopes, understanding the clonal structure of the tumor is essential in vaccine design. Clonality analysis can be done by targeted amplicon sequencing of selected mutations at the single cell level. Based on clonality analysis, epitopes resulting from variants that are expressed in the majority of tumor cells are better candidates.

Summary

Cancer cells acquire a phenotype of genomic instability. It may facilitate the oncogenesis as well as provide abundant mutational substrates for immune recognition. Neoantigens derived from the mutations are now routinely identified and being tested in clinical trials. This new approach for personalized immunotherapy is heavily dependent on the bioinformatics pipelines to map the reads, call the mutations and predict the immunogenicity of neoepitopes harboring the mutations. The bioinformatics pipelines may also need to be optimized for different sequencing platform. It is conceivable to perform both exome sequencing and RNA-Seq to be more confident about the mutation calling. While it is still challenging to predict tumor-rejecting neoepitopes as well as monitor immune responses to them, it is an exciting time to test personalized cancer immunotherapy by itself as well in combination with other therapies.

Acknowledgements

We acknowledge Dr. Pramod Srivastava for his inspiration, guidance and vision for this article and neoepitope research in general. Having been working on neoepitopes for over 30 years, Dr. Srivastava is a living legend who believes that the only cancer-specific epitopes are those epitopes derived from random passenger mutations from dividing cancerous cells.

Compliance with ethics guidelines

Sahar Al Seesi, Alok Das Mohapatra, Arpita Pawashe, Ion I. Mandoiu, and Fei Duan declare that they have no conflict of interest. This manuscript is a review article and does not involve a research protocol requiring approval by the relevant institutional review board or ethics committee.

References

Al Seesi S, Tiagueu Y T, Zelikovsky A, Măndoiu I I (2014). Bootstrap-based differential gene expression analysis for RNA-Seq data with and without replicates. BMC Genomics, 15(8 Suppl 8): S2

Basombrio M A (1970). Search for common antigenicities among twenty-five sarcomas induced by methylcholanthrene. Cancer Res, 30(10): 2458–262

Bentley D R, Balasubramanian S, Swerdlow H P, Smith G P, Milton J, Brown C G, Hall K P, Evers D J, Barnes C L, Bignell H R, Boutell J M, Bryant J, Carter R J, Keira Cheetham R, Cox A J, Ellis D J, Flatbush M R, Gormley N A, Humphray S J, Irving L J, Karbelashvili M S, Kirk S M, Li H, Liu X, Maisinger K S, Murray L J, Obradovic B, Ost T, Parkinson M L, Pratt M R, Rasolonjatovo I M, Reed M T, Rigatti R, Rodighiero C, Ross M T, Sabot A, Sankar S V, Scally A, Schroth G P, Smith M E, Smith V P, Spiridou A, Torrance P E, Tzonev S S, Vermaas E H, Walter K, Wu X, Zhang L, Alam M D, Anastasi C, Aniebo I C, Bailey D M, Bancarz I R, Banerjee S, Barbour S G, Baybayan P A, Benoit V A, Benson K F, Bevis C, Black P J, Boodhun A, Brennan J S, Bridgham J A, Brown R C, Brown A A, Buermann D H, Bundu A A, Burrows J C, Carter N P, Castillo N, Chiara E, Catenazzi MChang S, Neil Cooley R, Crake N R, Dada O O, Diakoumakos K D, Dominguez-Fernandez B, Earnshaw D J, Egbujor U C, Elmore D W, Etchin S S, Ewan M R, Fedurco M, Fraser L J, Fuentes Fajardo K V, Scott Furey W, George D, Gietzen K J, Goddard C P, Golda G S, Granieri P A, Green D E, Gustafson D L, Hansen N F, Harnish K, Haudenschild C D, Heyer N I, Hims M M, Ho J T, Horgan A M, Hoschler K, Hurwitz S, Ivanov D V, Johnson M Q, James T, Huw Jones T A, Kang G D, Kerelska T H, Kersey A D, Khrebtukova I, Kindwall A P, Kingsbury Z, Kokko-Gonzales P I, Kumar A, Laurent M A, Lawley C T, Lee S E, Lee X, Liao A K, Loch J A, Lok M, Luo S, Mammen R M, Martin J W, McCauley P G, McNitt P, Mehta P, Moon K W, Mullens J W, Newington T, Ning Z, Ling Ng B, Novo S M, O'Neill M J, Osborne M A, Osnowski A, Ostadan O, Paraschos L L, Pickering L, Pike A C, Pike A C, Chris Pinkard D, Pliskin D P, Podhasky J, Quijano V J, Raczy C, Rae V H, Rawlings S R, Chiva Rodriguez A, Roe P M, Rogers J, Rogert Bacigalupo M C, Romanov N, Romieu A, Roth R K, Rourke N J, Ruediger S T, Rusman E, Sanches-Kuiper R M, Schenker M R, Seoane J M, Shaw R J, Shiver M K, Short S W, Sizto N L, Sluis J P, Smith M A, Ernest Sohna Sohna J, Spence E J, Stevens K, Sutton N, Szajkowski L, Tregidgo C L, Turcatti G, Vandevondele S, Verhovsky Y, Virk S M, Wakelin S, Walcott G C, Wang J, Worsley G J, Yan J, Yau L, Zuerlein M, Rogers J, Mullikin J C, Hurles M E, McCooke N J, West J S, Oaks F L, Lundberg P L, Klenerman D, Durbin R, Smith A J (2008). Accurate whole human genome sequencing using reversible terminator chemistry. Nature, 456(7218): 53–59

Berman J N, Chiu P P L, Dellaire G (2014). Preclinical animal models for cancer genomics. In: Dellair G, Berman J N, Arceci R J, eds. Cancer Genomics: from Bench to Personalized Medicine, Elsevier Inc., 110–126

Bielas J H, Loeb K R, Rubin B P, True L D, Loeb L A (2006). From the Cover: Human cancers express a mutator phenotype. Proc Natl Acad Sci USA, 103(48):18238–18242

Blanchard T, Srivastava P K, Duan F (2013). Vaccines against advanced melanoma. Clin Dermatol, 31(2): 179–190

Boland J F, Chung C C, Roberson D, Mitchell J, Zhang X, Im K M, He J, Chanock S J, Yeager M, Dean M (2013). The new sequencer on the block: comparison of Life Technology's Proton sequencer to an Illumina HiSeq for whole-exome sequencing. Hum Genet, 132(10): 1153–1163

Bolger A M, Lohse M, Usadel B (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics, 30(15): 2114–2120

Boon T, van der Bruggen P (1996). Human tumor antigens recognized by T lymphocytes. J Exp Med, 183(3): 725–729

Castle J C, Kreiter S, Diekmann J, Löwer M, van de Roemer N, de Graaf J, Selmi A, Diken M, Boegel S, Paret C, Koslowski M, Kuhn A N,

Britten C M, Huber C, Türeci O, Sahin U (2012). Exploiting the mutanome for tumor vaccination. Cancer Res, 72(5): 1081–1091

Cheon D J, Orsulic S (2011). Mouse models of cancer. Annu Rev Pathol, 6(1): 95–119

Dranoff G (2012). Experimental mouse tumour models: what can be learnt about human cancer immunology? Nat Rev Immunol, 12(1): 61–66

Duan F, Duitama J, Al Seesi S, Ayres C M, Corcelli S A, Pawashe A P, Blanchard T, McMahon D, Sidney J, Sette A, Baker B M, Mandoiu I I, Srivastava P K (2014). Genomic and bioinformatic profiling of mutational neoepitopes reveals new rules to predict anticancer immunogenicity. J Exp Med, 211(11): 2231–2248

Duan F, Lin Y, Liu C, Engelhorn M E, Cohen A D, Curran M, Sakaguchi S, Merghoub T, Terzulli S, Wolchok J D, Houghton A N (2009). Immune rejection of mouse tumors expressing mutated self. Cancer Res, 69(8): 3545–3553

Duitama J, Srivastava P K, Măndoiu I I (2012). Towards accurate detection and genotyping of expressed variants from whole transcriptome sequencing data. BMC Genomics, 13(2 Suppl 2): S6

Feng J, Meyer C A, Wang Q, Liu J S, Shirley Liu X, Zhang Y (2012). GFOLD: a generalized fold change for ranking differentially expressed genes from RNA-seq data. Bioinformatics, 28(21): 2782–2788

Foley E J (1953). Antigenic properties of methylcholanthrene-induced tumors in mice of the strain of origin. Cancer Res, 13(12): 835–837

Gubin M M, Zhang X, Schuster H, Caron E, Ward J P, Noguchi T, Ivanova Y, Hundal J, Arthur C D, Krebber W J, Mulder G E, Toebes M, Vesely M D, Lam S S, Korman A J, Allison J P, Freeman G J, Sharpe A H, Pearce E L, Schumacher T N, Aebersold R, Rammensee H G, Melief C J, Mardis E R, Gillanders W E, Artyomov M N, Schreiber R D (2014). Checkpoint blockade cancer immunotherapy targets tumour-specific mutant antigens. Nature, 515(7528): 577–581

Kim D, Langmead B, Salzberg S L (2015). HISAT: a fast spliced aligner with low memory requirements. Nat Methods, 12(4): 357–360

Langmead B, Trapnell C, Pop M, Salzberg S L (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol, 10(3): R25

Larsen M V, Lundegaard C, Lamberth K, Buus S, Lund O, Nielsen M (2007). Large-scale validation of methods for cytotoxic T-lymphocyte epitope prediction. BMC Bioinformatics, 8(1): 424

Li B, Dewey C N (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics, 12(1): 323

Liu J, Blake S J, Smyth M J, Teng M W (2014). Improved mouse models to assess tumour immunity and irAEs after combination cancer immunotherapies. Clin Transl Immunology, 3(8): e22

Lundegaard C, Lund O, Nielsen M (2008). Accurate approximation method for prediction of class I MHC affinities for peptides of length 8, 10 and 11 using prediction tools trained on 9mers. Bioinformatics, 24(11): 1397–1398

Lurquin C, Van Pel A, Mariamé B, De Plaen E, Szikora J P, Janssens C, Reddehase M J, Lejeune J, Boon T (1989). Structure of the gene of tum- transplantation antigen P91A: the mutated exon encodes a peptide recognized with Ld by cytolytic T cells. Cell, 58(2): 293–303

Matsushita H, Vesely M D, Koboldt D C, Rickert C G, Uppaluri R, Magrini V J, Arthur C D, White J M, Chen Y S, Shea L K, Hundal J, Wendl M C, Demeter R, Wylie T, Allison J P, Smyth M J, Old L J, Mardis E R, Schreiber R D (2012). Cancer exome analysis reveals a T-cell-dependent mechanism of cancer immunoediting. Nature, 482 (7385): 400–404

McGranahan N, Furness A J, Rosenthal R, Ramskov S, Lyngaa R, Saini S K, Jamal-Hanjani M, Wilson G A, Birkbak N J, Hiley C T, Watkins T B, Shafi S, Murugaesu N, Mitter R, Akarca A U, Linares J, Marafioti T, Henry J Y, Van Allen E M, Miao D, Schilling B, Schadendorf D, Garraway L A, Makarov V, Rizvi N A, Snyder A, Hellmann M D, Merghoub T, Wolchok J D, Shukla S A, Wu C J, Peggs K S, Chan T A, Hadrup S R, Quezada S A, Swanton C (2016). Clonal neoantigens elicit T cell immunoreactivity and sensitivity to immune checkpoint blockade. Science, 351(6280): 1463–1469

McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo M A (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res, 20 (9): 1297–1303

Monach P A, Meredith S C, Siegel C T, Schreiber H (1995). A unique tumor antigen produced by a single amino acid substitution. Immunity, 2(1): 45–59

Nicolae M, Mangul S, Măndoiu I I, Zelikovsky A (2011). Estimation of alternative splicing isoform frequencies from RNA-Seq data. Algorithms Mol Biol, 6(1): 9

Noguchi Y, Chen Y T, Old L J (1994). A mouse mutant p53 product recognized by $CD4^+$ and $CD8^+$ T cells. Proc Natl Acad Sci USA, 91 (8): 3171–3175

Nowell P C (1976). The clonal evolution of tumor cell populations. Science, 194(4260): 23–28

Pandey V, Nutter R C, Prediger E (2008).Applied Biosystems SOLiD™ System: Ligation-Based Sequencing. In: Janitz M, ed. Next Generation Genome Sequencing. Wiley-VCH Verlag GmbH & Co. KGaA. p. 29–42

Prehn R T, Main J M (1957). Immunity to methylcholanthrene-induced sarcomas. J Natl Cancer Inst, 18(6): 769–778

Roberts A, Trapnell C, Donaghey J, Rinn J L, Pachter L (2011). Improving RNA-Seq expression estimates by correcting for fragment bias. Genome Biol, 12(3): R22

Robinson M D, McCarthy D J, Smyth G K (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics, 26(1): 139–140

Schmieder R, Edwards R (2011). Quality control and preprocessing of metagenomic datasets. Bioinformatics, 27(6): 863–864

Schuler M M, Nastke M D, Stevanovikć S (2007). SYFPEITHI: database for searching and T-cell epitope prediction. Methods Mol Biol, 409: 75–93

Srivastava P K (2015). Neoepitopes of Cancers: Looking Back, Looking Ahead. Cancer Immunol Res, 3(9): 969–977

Thomas R K, Nickerson E, Simons J F, Jänne P A, Tengs T, Yuza Y, Garraway L A, LaFramboise T, Lee J C, Shah K, O'Neill K, Sasaki H, Lindeman N, Wong K K, Borras A M, Gutmann E J, Dragnev K H, DeBiasi R, Chen T H, Glatt K A, Greulich H, Desany B, Lubeski C K, Brockman W, Alvarez P, Hutchison S K, Leamon J H, Ronan M T, Turenchalk G S, Egholm M, Sellers W R, Rothberg J M, Meyerson M (2006). Sensitive mutation detection in heterogeneous cancer specimens by massively parallel picoliter reactor sequencing. Nat Med, 12(7): 852–855

Tian S, Maile R, Collins E J, Frelinger J A (2007). CD8 + T cell activation is governed by TCR-peptide/MHC affinity, not dissociation rate. J Immunol, 179(5): 2952–2960

Trapnell C, Pachter L, Salzberg S L (2009). TopHat: discovering splice junctions with RNA-Seq. Bioinformatics, 25(9): 1105–1111

Yadav M, Jhunjhunwala S, Phung Q T, Lupardus P, Tanguay J, Bumbaca S, Franci C, Cheung T K, Fritsche J, Weinschenk T, Modrusan Z, Mellman I, Lill J R, Delamarre L (2014). Predicting immunogenic tumour mutations by combining mass spectrometry and exome sequencing. Nature, 515(7528): 572–576

Yates L R, Campbell P J (2012). Evolution of the cancer genome. Nat Rev Genet, 13(11): 795–806