Just how transformative will AI/ML be for immuno-oncology?

Daniel Bottomly, Shannon McWeeney

ABSTRACT
Immuno-oncology involves the study of approaches which harness the patient’s immune system to fight malignancies. Immuno-oncology, as with every other biomedical and clinical research field as well as clinical operations, is in the midst of technological revolutions, which vastly increase the amount of available data. Recent advances in artificial intelligence and machine learning (AI/ML) have received much attention in terms of their potential to harness available data to improve insights and outcomes in many areas including immuno-oncology. In this review, we discuss important aspects to consider when evaluating the potential impact of AI/ML applications in the clinic. We highlight four clinical/biomedical challenges relevant to immuno-oncology and how they may be able to be addressed by the latest advancements in AI/ML. These challenges include (1) efficiency in clinical workflows, (2) curation of high-quality image data, (3) finding, extracting and synthesizing text knowledge as well as addressing, and (4) small cohort size in immunotherapeutic evaluation cohorts. Finally, we outline how advancements in reinforcement and federated learning, as well as the development of best practices for ethical and unbiased data generation, are likely to drive future innovations.

INTRODUCTION
With the development of several breakthrough immunotherapies including anti-PD-1/PD-L1 and anti-CTLA-4 immune checkpoint inhibitors (ICIs) reviewed in Ribas and Wolchok as well as chimeric antigen receptor (CAR) T cells reviewed in June et al, immuno-oncology (IO) has been established as a promising framework for the development of cancer therapeutics. However, a number of challenges still remain. For instance, only a subset of patients with cancer diagnoses that would be otherwise terminal see durable clinical benefit from a given immunotherapy. This has led to increased interest in development of personalized biomarkers to ensure rational development of clinical trials. Of those patients who do respond, as is also the case with chemotherapy and targeted inhibitors, they can eventually develop resistance (as reviewed in ). As a means of circumventing resistance, much focus is currently placed on the development of ICI combinations, reviewed in Sanmamed et al as well as coupling an ICI with cytokines (or corresponding blocking inhibitors or antibodies) to fine tune the desired immune response as reviewed in Berraondo et al. In addition to resistance, the occurrence of adverse events including observations of “hyperprogressive” disease presents a challenge to the development of clinically viable therapeutics. It is increasingly recognized that the complexities in immunotherapeutic treatment are in part due to immunoeediting, reviewed in and . A major tenant of the immunoediting hypothesis describes how the immune system can be modulated by tumors to maintain a stable disease state or escape, allowing progression. These interactions between tumor-intrinsic pathways, tumor microenvironment (TME) and immune mechanisms produce substantial heterogeneity in tumor dynamics and therefore patient response. At the same time advances in computational power and storage as well as in the theory and practice of machine learning (ML) have accelerated our ability to gain insights from large multimodal datasets. ML approaches traditionally used for biomarker discovery or immune-therapeutic response have been reviewed previously.

This review focuses on recent advancements in artificial intelligence (AI) and ML, namely state-of-the-art neural network architectures and related applications. We note that there are numerous reviews that have focused on validation of IO targets and development of effective immunotherapies. Therefore, we have focused on other areas where AI/ML holds promise to improve IO in clinical care.

AI and ML are often used interchangeably and while related are still distinct concepts. AI as a field was originally started in the 1950s and focuses on computational approaches to mimic the ways that humans think in order to perform complex tasks. ML is a subfield of AI that uses algorithms trained on data to produce models that can perform complex tasks. We note that often use of the term AI is colloquially associated with a specific type.
of ML model known as a neural network. In particular, so-called deep neural networks (DNNs) have been shown to approach human level performance on a number of tasks. Traditionally, neural networks are conceptualized as having fully connected layers of nodes starting at the input with intermediary “hidden” layers terminating at the output layer (termed a feedforward network; figure 1). Not counting the input layer which is specified by the data, the depth is typically considered to be the number of remaining layers, while the width is measured as the number of nodes in a given layer. Although modern neural network layers can be composed of different types of components and connections, DNNs by definition all have multiple hidden layers. The power of DNNs is that they are thought to be able to represent complex relationships by encoding more general information in the early layers followed by increasing levels of abstraction. Arguably, to date, the most potentially impactful of these DNN technologies is the generative pretrained transformer version 4 (GPT-4) large language model (LLM) which has been shown to score respectably on a number of accreditation and academic exams. The primary interface to these models is a prompt-based question-and-answer framework tuned by reinforcement learning from human feedback (RLHF). In addition to the adaptation of GPT-4 into an “AI chatbot for medicine”, many of the concepts that make the GPT model successful can also be applied as a whole or in part to the study of IO (as well as other clinical research fields) as we discuss below.

It is critical that we can discern reality from the hype when it comes to AI/ML. Hype can distort our understanding of any new technology which makes it difficult to understand where it is best used. The Gartner hype cycle provides a graphical representation to represent the maturity, adoption, and social application of specific technologies. Recognizing that the time frame for AI/ML is not recent, the hype cycle can allow us to assess expectations versus reality for specific subsets of AI/ML such as DNNs or LLM. While there have been numerous criticisms of the hype cycle, it serves as a useful reminder that popularity of a technology may not often coincide with its maturity, leading to premature applications or misuse. The implications of this in the healthcare setting are far-reaching beyond economic impact. Therefore, we need to consider “how soon is now”, that is, which approaches will be realized in the near, mid and long term? In addition to the maturity of the technology, we need to also consider organization maturity and readiness for the development, deployment and maintenance of AI/ML. The term “Technical Debt”, originally defined in software engineering, has been used to describe trade-offs in the long-term costs of adopting AI/ML systems. Quickly deploying an AI/ML model often results in increased future expenses to maintain performance and stability when faced with changes to the underlying code, data format/distributions as well as computing or inference environment. To address these challenges, machine learning operations (MLOps) was devised as an
The remainder of the paper highlights four clinical/biomedical challenges relevant to IO, as well as other clinical research fields, and how current AI/ML methodologies may help in the near/mid-term timeframe (figure 3). Finally, we offer thoughts about key innovations and practices necessary for long-term impact.

Challenges of efficiency in clinical workflows

Although much attention is given to the application of DNNs and other advanced ML systems in biomedical and clinical research, clinical practice and administration stand to benefit as well and are likely a near-term beneficiary of advancements in AI/ML. For example, AI/ML medical scribes can automate documentation, improving accuracy, reducing physician burden, and enhancing patient care.

Another promising area is the ability to increase patient engagement and adherence to treatment. Approaches to increasing patient engagement have been systematically reviewed previously. Access to LLM chatbots as well as mobile apps are newer ways that can provide a patient greater access to their health information. A recent study found similar accuracy between humans and ChatGPT in answering questions related to clinical genetics. However, it was noted that ChatGPT displayed poorer performance if critical thought was required. One of the strengths of ChatGPT as well as other LLMs is the ability to rapidly summarize prompted information. This is an ability which could be readily adapted to generating discharge summaries and is seeing rapid uptake in medical transcription. However, as with the other more critical diagnostic and treatment uses, use of specific LLMs for these purposes would have to be rigorously validated. The successful application and adoption of DNN powered applications have great potential for lowering the cost of delivering healthcare. A recent study found that utilization of DNNs for treatment in particular could potentially result in large cost savings for hospitals. Another challenge for healthcare systems is the management of adverse drug reactions, estimated to cost US$30.1 billion annually in the USA alone. Data on adverse events are gathered as part of “pharmacovigilance” efforts and AL/ML approaches are considered to be a tool to help with case processing and assessment or as part of “human-in-the-loop” systems. AI/ML systems designed to predict adverse events ahead of time such as the adverse events atlas are under development and have the potential to improve patient care as well as realize significant cost savings.

Adverse event detection extends beyond the pharmacological use cases to other aspects of patient risk. An area of clear interest is AI/ML prediction of point-of-care risk stratification which has received tremendous attention.
and already provided key use cases on the challenges of implementation. External validation of a proprietary sepsis prediction model which had been implemented in hundreds of hospitals found that the model predicted the onset of sepsis with an area under the curve of 0.63, which is substantially worse than the performance reported by the vendor, highlighting the need for external validation before adoption.38 With regard to oncology and immunology, applications have included quality improvement by identification of oncology patients at risk for a near-term emergency room visit,39 prediction of mortality in immunotherapy patients40 and early identification of patients with who could realize durable clinical benefit from PD-(L)1 blockade-based ICI treatment and chemotherapy,41 among others. Best practices for implementing risk stratification models in the clinic are actively evolving—one common theme is that use of a “plug and play” model is fraught with issues given concerns about clinical relevance of predictions, the lack of considerations for how it will be integrated into workflows, as well as the need for training and change management.42–45 Evaluation of the models must be intuitive to clinicians and ideally focused on metrics related to quality of care and patient outcomes rather than technical performance of the model.46 As with all clinical decision models, we must also address ethical concerns such as bias and fairness as it relates to both the models and the data the models were trained on. While we view many aspects of AI/ML improving efficiency in clinical workflows as candidates for near term benefit, for the reasons above, the risk stratification approaches are likely more near to mid-term.

For academic medical centers and cancer centers, the impact of the savings due to increased efficiency from the near-term applications could potentially be used not only for expansion of new clinics, but also implementation of immunotherapeutic trials and other clinical and translational research innovation.

Challenge of curation of high-quality image data
The advent of digital pathology, reviewed in Baxi et al,47 has led to an increasing reliance on AI/ML methodology and is a near term beneficiary of advancements in AI/ML. The importance of digital pathology for IO should not be overlooked (see Morad et al48 and references therein). The automation of tissue slides in digital pathology, in conjunction with the rapid advances in multiplex imaging
technology, sets the stage for extensive characterization of the TME.

Relevant for digital pathology, we provide additional background on ML. Traditionally, ML approaches are classified into two main types: supervised and unsupervised. Supervised learning requires the presence of a high-quality “label”, often a classification or outcome. In the case of clinical research, this typically requires either collection of requisite data and evaluation of patient response, say to a given immunotherapeutic, by clinicians or access to existing collections of well-curated and annotated images. The data collection and evaluation process can be time-consuming and laborious and often limits the utility of retrospective data. Unsupervised learning on the other hand does not require prelabeled data, making it more appealing for discovery in clinical research such as for new biomarkers. However, datasets used for biomarker discovery in particular tend to have large numbers of correlated features. In addition, genomics data are also often impacted by unknown technical factors, reviewed in Leek et al. These issues make true unsupervised biomarker discovery especially challenging regarding validation and interpretability. However, there are other approaches to model training such as reinforcement learning (discussed in the Future Impact and Concluding Remarks section) and hybrid approaches of supervised and unsupervised learning which provide a bridge between a priori knowledge and discovery.

In digital pathology, while high quality classification/ outcome/annotation information exists, often there are large stores of unannotated images are also available. Semisupervised approaches seek to leverage both characterized and uncharacterized data to improve model performance. Related to semisupervised learning is self-supervision reviewed in Krishnan et al. This type of training formulates a supervised learning problem based on a secondary outcome that can be derived from the data itself. Although this conceivably can be done in many ways, two common approaches exist: contrastive learning and generative (masked) learning. Contrastive learning first develops a model that can distinguish, for example, whether two images are similar or different. In this manner, it captures important features for distinguishing images without explicit labeling. Masked learning on the other hand removes a portion of the data, says a word in a sentence, and requires the model to predict the word. The process guides the model to infer approximations of the underlying context.

Another approach to dealing with limited curated data for algorithm development is to instead generate more high-quality data in silico. The ability to generate new realistic data and content is a powerful and controversial development in AI/ML. These so-called generative neural network architectures are commonly used for image, speech and text generation. The fundamental type of generative neural network is the Generative Adversarial Network (GAN). GANs consist of two interlocked models, a generator, which models the data distribution producing a synthesized result and a discriminator which tries to distinguish the synthesized version from the training data. GANs have been developed to help with digital pathology, analyzing histological images of tumor samples both in terms of generating new high-quality images but also in learning latent representations in an unsupervised manner. Additional uses in pathology include preprocessing, color normalization, virtual staining, image enhancement, removal of ink marks and augmentation) as well as other types of analyses including nuclei detection, segmentation and domain adaptation. Generative neural network approaches have also been developed for the analysis of single-cell data in general including prediction of immune response from a baseline measurement. Of high relevance to IO are spatial transcriptomics technologies that combine gene expression (including single-cell) and imaging. Similar to imaging, there is much interest in the development of AI/ML analysis approaches. For instance, the recently proposed GraphST method leverages graph neural networks refined using contrastive learning.

**Challenge of finding, extracting and synthesizing knowledge**

Electronic health records (EHRs) have become an increasingly important source of real-world data (RWD). In turn, RWD can enhance the development of IO therapeutics and companion diagnostics as well as provide critical real-world evidence for regulatory approval. Algorithmic and technical innovations coupled with access to huge quantities of text data from the internet enabled the creation of LLMs such as Bidirectional Encoder Representations from Transformers (BERT) and the GPT. These models can have billions of parameters, for instance, the previous generation GPT-3 had 175 billion, which encode the capacity for highly sophisticated pattern recognition and generalization. Because of their potential they are considered to be “foundation models” with utility in many different tasks. In the case of GPT-3, it was observed that “few-shot learning”, in this case applying in-line text examples for the system, produced performance on par with many state-of-the-art systems that had been specially tuned using labeled training data. This feature has been explored in the clinical domain. Additionally, LLMs specific to science and biomedical tasks have been created which may be better suited for technical use-cases. In fact, it was observed that using only biomedical texts in pretraining provided a benefit over models that included other less relevant sources of information. In addition to the general clinical applications reviewed in, of most relevance to IO research, is the ability to retrieve and summarize existing research. One of the challenges facing widespread adoption of methods in this area is the observation that LLMs can “hallucinate” producing realistic sounding content that is factually incorrect. As mentioned on their homepage, the prevalence of hallucination led to one scientific LLM endeavor, Galactica, removing its demo from the web. This is due to these models encoding knowledge as a
function of their learnt weights—not necessarily facts. Ongoing work is exploring how to incorporate prior knowledge in the form of graphs which may be able to help address this issue.

The ChatGPT model in particular is incredibly powerful, however, it can still produce inaccurate results and therefore caution is needed. For instance, in online supplemental figure S1A, an interaction is shown where ChatGPT is asked to list approved immunotherapies, their indications, companion diagnostics and references. Although, it left blank most of the indications, generally the companion diagnostic information seemed accurate with the exception of CD19 expression for Axicabtagene ciloleucel CAR-T therapy. One possible source of ambiguity is that Axicabtagene ciloleucel is an anti-CD19 CAR-T therapy, although having a positive expression of CD19 is not a prerequisite for treatment. To attempt to gain further insight into what chatGPT classifies as a companion diagnostic, it was further asked to define a companion diagnostic. The chatGPT system described the companion diagnostic as “a test or assay used alongside a specific drug to identify patients who will benefit from the treatment” (online supplemental figure S1C). This highlights that there are still nuances not recognized by the LLMs. One way to combat hallucinations as well as other ambiguous responses is to ask the model to check itself for accuracy. Finally, we provided the above definition of companion diagnostic and asked it to confirm if CD19 is considered a companion diagnostic for Axicabtagene ciloleucel. Again, the system answered in the affirmative justifying its reasoning (online supplemental figure S1D). This further emphasizes the need for conversational interactions with the system as opposed to strict question and answer. Similarly, in online supplemental figure S1B, when asked to verify the references, the system did indicate that the references were placeholders. Hopefully in the future, if correct references cannot be given these fields can be left blank or at least clarified to the user as part of the response text to avoid potential issues.

LLMs can also be used to extract unstructured or semi-structured information found in EHRs as is reviewed in Fu et al. This is highly relevant to IO as a means to gather data that would otherwise be laborious to achieve, for example, tumor-infiltrating lymphocyte classification. Additionally, EHR information can be queried to evaluate patients with respect to clinical trial eligibility. In a recent adaption of this approach, using CT-BERT, a version of BERT fine-tuned on ClinicalTrials.gov data an AI/ML framework was proposed to assess the generalizability of a given clinical trial with respect to its eligibility criteria. The recently developed LLM GatorTron focused not only on computational phenotyping/cohort characterization but also on its potential use in pharmacovigilance. One of the current challenges with application of ChatGPT to EHR mining tasks concerns patient privacy. This is because currently interactions with ChatGPT are transmitted over the internet. Additionally, if limited to basic (non-identifying) diagnostic information, the performance of ChatGPT degrades considerably. These and related issues have influenced the development of LLMs derived specifically from deidentified clinical notes as well as other relevant sources. Despite general advances in LLM capabilities, the robust application to biomedical and clinical text is likely a mid-term beneficiary of AI/ML advancements. We note that implementation of the largest scale LLMs is often beyond the resource of individual cancer centers so optimization and adaptation of existing models will likely be key.

Challenge of small cohort size in immunotherapeutic evaluation cohorts

One of the challenges with developing treatment strategies or evaluating biomarkers for personalized therapy is that most clinical patient cohorts tend to be small and heterogeneous. This more often than not leads to the lack of novel discovery and/or generalizability.

An often-pursued approach especially in biomarker studies is to develop models first in larger cohorts such as The Cancer Genome Atlas (TCGA) and then apply them to more focused studies. For instance, using a curated list of 29 functional gene expression signatures describing TME processes, 4 signatures related to melanoma were derived using TCGA that could predict response to ICI in smaller independent studies.

The concept of “pretraining” larger DNN models is commonly used in other fields, often with the goal of leveraging these models (termed sources) to help solve other problems similar in form to the original task (termed targets), a procedure referred to as transfer learning. With a large pretrained model in hand, transfer learning can be implemented using a number of different strategies. This approach relies on the assumption that earlier layers in a large DNN have learnt sufficiently general patterns to be useful to other tasks. Learnt weights within the source DNN model are either “frozen” or used as the starting point for additional fine-tuning with the target dataset as reviewed in the study of medical image classification. In addition to imaging, with the advent of single-cell transcriptomics, large quantities of publicly available data can be collected to pretrain models for the purpose of transfer learning. One variation of this is annotation of single-cell experiments relative to existing datasets. As larger single-cell atlases are generated such as the human cell landscape, related approaches can be used to map single-cell datasets onto the atlases to facilitate analyses.

Based on the success of models such as GPT and BERT, there is much interest in developing large pretrained transformer models to support transfer learning in many aspects of clinical research including IO. The transformer architecture was originally developed in the context of sequence (eg, language) modeling using DNNs. At its core is the concept of self-attention which was originally devised as a means to reduce computational complexity, increase parallelizability and to improve detection of long-range dependences in these sequence models. However, importantly, at the same time it also provides a means of...
model interpretation. Interestingly, in addition to the utilization of the pretrained transformer architecture in state-of-the-art LLMs, models have also been devised to pretrain on available large biomedical datasets. Using publicly available single-cell data, pretrained transformer models have been devised to capture network dynamics such as Geneformer\textsuperscript{95} as well as begin to build more foundational generative models such as scGPT\textsuperscript{94} and tGPT.\textsuperscript{95} The tGPT model in particular showed promise for differentiating immunotherapeutic treatment outcomes in urothelial carcinoma leveraging available bulk RNASeq data. In addition to single-cell data, models are being developed that utilize clinical and mutational data for ICI treatment outcome prediction such as the Clinical Transformer.\textsuperscript{96} Given challenges with relevant dataset accrual and potential issues in bias, it is likely this is a mid-term to long-term beneficiary of AI/ML in the clinic.

Future impact and concluding remarks
Distinct from more traditional supervised learning described previously, reinforcement learning provides a mechanism through which models can be continuously trained. This is accomplished through the accumulation of rewards from chosen actions with the goal being to maximize the total received rewards in the long term. This approach, though currently used infrequently in biomedical research, has been implemented to predict drug sensitivity as well as optimizing chemotherapeutic and radiotherapy doses in retrospective and simulated clinical settings as reviewed in Eckardt \textit{et al.}\textsuperscript{95 98} Similar approaches are also being adapted to assess immunotherapeutic challenges such as achieving control of the balance of a patient’s immune and tumor cells with respect to treatment.\textsuperscript{99} Use of reinforcement learning has also been successfully used in the search for T cell receptor beta chain CDR3 sequences that have enhanced affinity for peptide sequences.\textsuperscript{100} This has implications for adoptive T cell immunotherapy. One breakthrough in the application of reinforcement learning has been the successful use as a means to fine-tune the interaction with a given LLM through RLHF\textsuperscript{101} as in ChatGPT. With the accelerated adoption of large DNNs in IO research variations on this approach have the potential to allow the research community as well as key stakeholders such as clinicians and other subject matter experts to be able to refine predictions of these models in an efficient manner.

One of the main challenges facing deployment of DNN systems in healthcare and biomedical research today is access to data. State-of-the-art DNN models with potentially billions of parameters need proportionally large datasets which may be more than a single institution or organization may have. At the same time concerns with privacy, security, computational constraints as well as intellectual property can limit use of a shared repository. As reviewed in Rieke \textit{et al.}\textsuperscript{102} the implementation of digital healthcare can use advances in federated learning to help address these concerns. Federated learning allows the sharing of DNN model optimization updates and parameters in order to learn a consensus model that performs better than models trained in isolation. This approach can further be leveraged in the biomedical sciences as reviewed in Kaisiss \textit{et al.}\textsuperscript{103} For both healthcare and research, the use of federated learning alone is not guaranteed to address privacy and security concerns.

Aggregating data for secondary analysis/modeling can be challenging depending on the data types and sources. While AI/ML algorithms are often robust to individual issues (eg, different variable definitions/coding, mismatched distributions, diverse data types, missing data and class imbalance), biomedical data often is characterized as having many of these issues simultaneously.\textsuperscript{104} AI/ML methods can be key as part of data cleaning and quality assurance (QA) / quality control (QC) to detect data anomalies and quality issues, as well as to facilitate mapping and transformation for integration. One such approach for automated concept mapping used DNNs to match data and coding at individual institutions to those of the observational medical outcomes partnership common data model.\textsuperscript{105} Additionally, work carried out as part of the INCISIVE project provides a promising framework for unifying, harmonizing, and securely sharing scattered cancer-related data to aggregate the large datasets needed to develop and evaluate trustworthy AI models.\textsuperscript{106}

While the amount of data is clearly a critical factor, issues with bias and fairness have reiterated that the greatest potential for impact is when the data are generated. This has led to efforts focused on developing the tools and approaches to support prospective data collection. For example, the National Institutes of Health (NIH) Common Fund Bridge2AI project\textsuperscript{107} is focused on generating flagship biomedical data sets that are ethically sourced, well curated and accessible. The adoption of data generation best practices from these efforts in the key cancer consortia such as the Cancer Moonshot Immuno-oncology translational network\textsuperscript{108} and others will provide invaluable data for realizing the potential of AI/ML in IO.

Acknowledgements The authors thank Guest Editor Dr. Alan Hutson and the Journal for ImmunoTherapy of Cancer for the invitation to contribute to this series.

Contributors DB and SM jointly developed the outline for the paper. DB wrote the initial draft. SM provided additional text and references, as well as final editing. All authors read and approved the final manuscript.

Funding This study was funded by NCI 2U54CA224019-05 and NIH Common Fund 3OT2OD032644-01S1.

Competing interests None declared.

Patient consent for publication Not applicable.

Provenance and peer review Commissioned; externally peer reviewed.

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.


**Figure S1**

**(A)**

**Prompt:** List current FDA approved immunotherapies with indications, companion diagnostics and relevant citations in HTML table format

**chatGPT:** Certainly! Here’s a list of current FDA-approved immunotherapies along with their indications, companion diagnostics, and corresponding citations in an HTML table format:

<table>
<thead>
<tr>
<th>Immunotherapy</th>
<th>Indication</th>
<th>Companion Diagnostics</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pembrolizumab (Keytruda)</td>
<td>Multiple indications</td>
<td>PD-L1 expression (22C3 pharmDx assay)</td>
<td>PMID: 27659769</td>
</tr>
<tr>
<td></td>
<td>including:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nivolumab (Opdivo)</td>
<td>Multiple indications</td>
<td>PD-L1 expression (Dako PD-L1 IHC 28-8 pharmDx assay)</td>
<td>PMID: 26637106</td>
</tr>
<tr>
<td></td>
<td>including:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Atezolizumab (Tecentriq)</td>
<td>Multiple indications</td>
<td>PD-L1 expression (Ventana PD-L1 SP142 assay)</td>
<td>PMID: 27087445</td>
</tr>
<tr>
<td></td>
<td>including:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ipilimumab (Yervoy)</td>
<td>Metastatic melanoma</td>
<td>N/A</td>
<td>PMID: 20525992</td>
</tr>
<tr>
<td>Axicabtagene ciloleucel</td>
<td>Diffuse large B-cell</td>
<td>CD19 expression (Flow cytometry)</td>
<td>PMID: 29768143</td>
</tr>
<tr>
<td>(Yescarta)</td>
<td>lymphoma</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**chatGPT:** Please note that this is a small sample of FDA-approved immunotherapies, and there are many more available. The provided table can serve as a starting point, and you can expand it with additional entries as needed.

**(B)**

**Prompt:** Please verify that the above citations correspond to the specific Immunotherapies

**chatGPT:** I apologize for the confusion in the previous response. Unfortunately, as an AI language model, I don't have direct access to external sources such as PubMed to verify specific citations in real-time. The citations provided in the previous response were placeholders and not specific to the immunotherapies mentioned. To obtain accurate and up-to-date information on FDA-approved immunotherapies and their associated citations, I recommend referring to trusted sources such as the U.S. Food and Drug
How do you differentiate a drug target from a companion diagnostic? Summarize in 100 words or less

chatGPT: A drug target is a molecule, often a protein or nucleic acid, involved in a disease process that a drug aims to interact with for therapeutic effects. It is crucial for drug development. On the other hand, a companion diagnostic is a test or assay used alongside a specific drug to identify patients who will benefit from the treatment. The diagnostic test detects biomarkers or genetic mutations associated with the drug target. It ensures personalized treatment, helping physicians select patients likely to respond positively, reducing unnecessary treatments, and potential side effects while advancing personalized medicine and improving patient outcomes.

Is CD19 expression a companion diagnostic for Axicabtagene ciloleucel

chatGPT: Yes, CD19 expression can be considered a companion diagnostic for Axicabtagene ciloleucel (Yescarta). CD19 is a protein expressed on the surface of certain cancer cells, including B-cell lymphomas and leukemias. Axicabtagene ciloleucel is a CAR-T cell therapy that specifically targets CD19-expressing cancer cells.

As a companion diagnostic, testing for CD19 expression helps identify patients who are suitable candidates for treatment with Axicabtagene ciloleucel. Patients with CD19-positive B-cell malignancies are more likely to benefit from this CAR-T therapy because the treatment works by genetically modifying the patient's T cells to target and destroy CD19-expressing cancer cells.

Before undergoing Axicabtagene ciloleucel treatment, patients are typically tested for CD19 expression in their cancer cells to ensure that the therapy is appropriate for their specific cancer type. If CD19 is not expressed in sufficient levels, the treatment may not be effective, and alternative therapies may be considered.
It's important to note that medical information can evolve, and the status of companion diagnostics and treatment guidelines may change over time. Therefore, it's always best to consult with healthcare professionals and refer to the latest information from regulatory agencies and official sources for the most current and accurate details.

Figure S1. Interaction with chatGPT regarding FDA approved Immunotherapeutics.
Prompts are given as written with chatGPT’s response in grey. (A) and (B) are from interactions on June 6th 2023 on https://chat.openai.com/, May 24 version. While (C) is from July 26th 2023 and (D) is from July 27th 2003, July 20 version.